

DARPA dreaming

Replicating the success of the US Defense Advanced Research Projects Agency (DARPA), in an organization devoted to energy research, will be easier said than done.

Everyone always thought DARPA was cool. Last month, in a major study on US competitiveness, the National Academies suggested that the federal government build a new one — ARPA-E — to address energy research.

But what exactly is DARPA — and is it really that special? And if it has a magic all of its own, could it be replicated in a different time and place, when confronting different challenges?

The answer is complex. There's more to the venerable Pentagon research agency than meets the eye. Notwithstanding various Hollywood depictions of DARPA, it has never had any labs or opulent premises of its own. What marks it out, instead, are subtle structural touches that any successful imitator would need to recreate.

Three or four facets of DARPA set it apart. One is the loyal patronage of a leader — President Dwight D. Eisenhower when it was set up in response to Sputnik, and presidents and defence secretaries ever since. Without this support from the top, DARPA would have been extinguished by suspicious rivals in the army, navy and air force.

Second, at least in its hey-day, DARPA was not 'mission-driven' in the manner of, say, the National Institutes of Health. Most people probably think DARPA's role was to meet the needs of the army, navy and air force, but nothing could be further from the truth. The armed forces had their own labs and programmes to do that. DARPA spun out ideas that the forces said they didn't want, or hadn't even thought of. Defence secretaries used it as an agent for the type of sea change that the rest of the Pentagon could be relied on to resist.

Third, the agency has no bureaucracy or infrastructure to speak of. Its annual budget of \$3 billion is handled by a director, a deputy director, a handful of office chiefs and a few dozen programme directors, most of them on short tenure.

It does, however, operate an effective congressional liaison office. It is true that some of its work goes ahead without the usual scrutiny because it is secret, but most of it is open and subject to the usual oversight. The committee structure that oversees the Department of Defense is relatively simple, however, and a few champions on Capitol Hill can protect DARPA from meddling. Even then, some observers see a steady grinding down of the agency's soul. They say

it is getting more like the National Science Foundation, more "fair". Everyone likes to be fair. But for DARPA, what counts is being agile.

That agility has brought remarkable success over half a century. DARPA concepts led directly to military innovations such as stealth materials and pilotless aircraft, helping to win the cold war. At the same time, it openly conducted pioneering public projects such as Arpanet, which grew into the Internet (apologies to CERN).

Some dissenters — who are given space in the academies' report, *Rising Above the Gathering Storm* — complain about government picking winners, and some even claim that the Department of Energy's sprawling network of laboratories has done just as well as DARPA, dollar for dollar.

But DARPA's track record of success fully justifies the National Academies' call for an ARPA-E. Unfortunately, the academies' report is silent on the obstacles that would need to be surmounted for such a body to work.

The inside of ARPA-E would be the easy part — smart people recruited at high wages for short periods, backing whatever horse they fancy and cajoling their grantees to push the envelope harder while collaborating intensively. It's the exterior linkages that make the project hard.

The original DARPA had the iron-clad commitment of the defence secretary, the president and Congress. But energy secretaries are marginal figures in the federal government, and presidents may or may not find the time to pay attention. Congressional oversight of the Department of Energy, meanwhile, is a basket-case of greedy and conflicting interests.

ARPA-E is unlikely to fly in the way the academies suggests, unless the energy department is rebuilt from top to bottom. But in different contexts — other nations, for example, facing other challenges — the lessons of DARPA's success are there to be learnt. Their resonance can only grow as research agencies around the world get larger, more comfortable, more audited and more risk-averse. ■

"ARPA-E is unlikely to fly in the way the National Academies suggests, unless the energy department is rebuilt from top to bottom."

A less toxic solution

Industry should get behind a European partnership that will explore alternatives to animal testing.

A public-private partnership established by the European Commission this week will boost the development of alternative methods to the animal testing of chemicals. More than 10 million animals are used each year in Europe to test chemicals for

safety. Now Europe is getting serious about developing alternative approaches (see page 144). Chemical manufacturers and political leaders have joined the animal lobby in embracing the alternatives, partly because of the sheer cost of using animal tests to meet new chemical safety requirements.

The European Commission's enterprise directorate this week hosted a conference on these alternatives, jauntily entitled 'Europe Goes Alternative'. It has taken three years of delicate negotiations to get industry on board, but at the meeting six trade associations representing hundreds of companies signed up to a Commission-led



project with the stodgier name of the European Partnership to Promote Alternative Approaches to Animal Testing. More signatures are expected shortly.

The text of the partnership agreement is rather bland, merely committing companies to agree that reduced use of animals in safety testing is a good idea. But it also commits the signatories to develop an action programme aimed at developing alternative methods. The Commission wants this action plan, which will be based on the sharing of information and the joint development of new approaches to testing strategies, to be in place by spring 2006.

It will need to be. Barring last-minute delays, controversial legislation on chemical testing will get its first reading in the European Parliament this week. The proposed Registration, Evaluation and Authorization of Chemicals (REACH) law would require regulatory approval for all chemicals sold in Europe — including some 30,000 compounds that have been around so long that they've never been registered before. Tests that do not require animals might greatly reduce the costs to industry of obtaining approval.

Scientists at the European Centre for the Validation of Alternative Methods (ECVAM) in northern Italy — which was set up by the European Commission to develop alternatives to animal testing — argue that animal tests are badly flawed. They say the new drive for alternative methods will improve the science of toxicity testing. And public safety demands that the new tests are shown to be better predictors of toxicity than the existing methods.

To this end, ECVAM scientists want chemicals manufacturers to provide more information, including data on compounds that have

been tested but not brought to market. Companies are reluctant to share this information for proprietary reasons. But it should be possible to derive shielding arrangements that will enable outside toxicologists to access it, without the release of commercially sensitive information about the products that were tested.

The action plan also calls for the sharing of the compounds themselves. These could be used to compare the efficiency of a new test against existing animal tests. It took ECVAM nearly a year to gather enough compounds to prove the value of its new *in vitro* skin irritation test, for example. The action plan would lead to simple procedures for material transfer that respect industry's concerns over proprietary information.

Perhaps the most difficult point in the action plan concerns its call for the release of more information on the performance of animal tests: how robust, reproducible and relevant are they? The data so far give grounds for concern. Yet industry has been resistant to this.

If the gold standard of animal tests against which new tests are to be compared turns out to be made of tin, the political fallout would be considerable. Public trust in the ability of regulatory authorities and industry to address safety issues would be damaged. But in the interests of a thorough, economically viable and scientifically valid product-safety testing regime, information about the methods used in the past needs to be shared, and fairly investigated. ■

"Public safety demands that the new tests be shown to be better predictors of toxicity than existing methods."

Flu in circulation

An interim US rule on safeguards may not, on its own, be enough to contain the 1918 flu virus.

The US Centers for Disease Control and Prevention (CDC) has published an interim rule placing the reconstructed 1918 flu virus on its list of select agents, and outlining provisions for its safe handling. But these are just the first steps that need to be taken to assure the public that the virus is in safe hands.

The interim rule, which was published in the *Federal Register* on 20 October, means that the virus may be shared with laboratories in the United States that have registered with the agency (see page 134). Some sharing is needed to accelerate progress in understanding its virulence — but it will also increase the risks of an accidental release. The classification of the virus is welcome, although some virologists would argue that it is overdue, given that the existence of the strain was well known months in advance of its publication (T. M. Tumpey *et al. Science* **310**, 77–80; 2005).

The 1918 flu virus is hard to contain and is capable of spreading rapidly between people. The researchers who work with the reconstructed virus point out that current flu vaccines and drugs provide good protection from it — but these are in short supply, and the threat of an accidental release is real.

The risks of such release during the physical shipping of the virus will be reduced if laboratories choose to construct it themselves, on

the basis of the published sequence. But that still leaves the risk of an escape from labs that work with it.

The CDC has ruled that enhanced biosafety level 3 laboratories can work with the virus, rejecting calls for a tougher, level-4 requirement that would have restricted the work to a handful of laboratories. That decision seems justifiable, in the interests of rapid research.

But uncertainty continues to cloud the question of access to the virus for laboratories abroad, where the CDC's writ doesn't run. Already, a biosafety level 4 lab in Winnipeg, Canada, has announced plans to reconstruct the virus.

No one will question the motives or the security arrangements at the Canadian lab, but the question of international regulation for this and other reconstructed viruses remains fraught. There is no international regime for the mandatory regulation of virus reconstruction, and it is hard to imagine how one could be put together in the time available.

In 1994, however, the World Health Organization (WHO) brokered an agreement restricting the smallpox virus to just two laboratories across the world. National governments should ask the WHO to examine the need for a broader agreement between member states to oversee the distribution of potentially dangerous, reconstructed viruses such as 1918 flu. ■

"The CDC has ruled that enhanced biosafety level 3 laboratories can work with the virus. That decision seems justifiable, in the interests of rapid research."

RESEARCH HIGHLIGHTS

Seafloor snacks

Biol. Lett. doi:10.1098/rsbl.2005.0374

Fossilized sea turtles from the extinct family Protostegidae have been found with bellies full of shellfish. The broken shells in the 110-million-year-old turtles' guts reveal that these ancient sea creatures fed on bottom-dwelling bivalves.

This is surprising, reports Benjamin Kear of the University of Adelaide, who discovered some of the fossils, because protostegids were thought to be surface feeders. Some of their modern relatives, including *Caretta* (pictured), eat mussel-like shellfish from the sea floor. But protostegids had primitive limbs thought to be unsuited to diving. One solution to this puzzle might be that they lived in shallow water. The fossils were found in the Toolebuc Formation in Queensland, Australia.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

H. HALL/OSF

GEOPHYSICS

After the earthquake

Geophys. Res. Lett. **32**, L20807 (2005)

The Sumatra earthquake last December not only sent a tsunami rushing across the sea, but also shot sound waves into space. This allowed a remarkable observation to be made — how a natural terrestrial phenomenon affects the magnetic field surrounding Earth.

By comparing data from Thailand, China and Japan, Kyoto University's Toshihiko Iyemori and his colleagues found localized long-wavelength pulsations in the geomagnetic field at the Thai city of Phimai. The timing and nature of the pulsations suggest that the magnetic disturbance was caused by sound waves from the earthquake.

Iyemori has also found that the massive 1991 Pinatubo volcanic eruption in the Philippines affected the magnetosphere.

DRUG DISCOVERY

Immune booster

Nature Chem. Bio. doi:10.1038/nchembio746 (2005)

By combining knowledge of chemical structure with an understanding of biological function, researchers have designed simple molecules that bind to a tumour necrosis factor (TNF) receptor. Such receptors help to control immune responses, so TNF impostor molecules could form the basis of therapies to boost or suppress the immune system.

Sylvie Fournel and Gilles Guichard, both at the Institute of Cellular and Molecular Biology in Strasbourg, France, and their colleagues made the small molecules by synthesizing the parts of a protein known as CD40L that bind it to its TNF receptor. They

attached these protein parts to a central scaffold that had the same threefold symmetry as the protein's receptor.

MICROBIOLOGY

Cells come unstuck

J. Exp. Med. **202**, 1–13 (2005)

In vitro experiments are unravelling details of how the bacterium *Helicobacter pylori* (pictured) causes gastric ulcers, the discovery that won this year's medicine Nobel Prize.

A protein produced by *H. pylori*, called CagA, is known to disrupt the protective lining of the stomach. Researchers led by Chihiro Sasakawa, of the University of Tokyo, show that it does this by binding to Crk proteins in the host cells. The hijacking of the Crk proteins activates signalling pathways that control cell turnover. This ultimately leads to the loss of the adhesion molecules that help the cells of the stomach lining to stick together, causing cell scattering.

CELL BIOLOGY

Waste not, want not

Cell **123**, 423–436 (2005)

The cell's waste-disposal machine, the proteasome, has hidden talents. As well as degrading unwanted proteins, this enzyme has a key role in controlling the activity of certain genes, say US researchers.

A team led by William Tansey, of the Cold Spring Harbor Laboratory in New York, and Jerry Workman, at the Stowers Institute for Medical Research in Kansas City, Missouri, has studied in yeast one of the two main parts that make up this complex enzyme. They found that it draws another collection of proteins — known as the SAGA coactivator complex — to the control regions of certain genes. This complex alters the chemistry of the proteins that package DNA, so altering gene activity.

STRUCTURAL BIOLOGY

Engines of creation

Science **310**, 827–834 (2005)

Of all the celebrated molecular machinery in biology, the most vital is surely the ribosome. This structure turns the genetic data of RNA into proteins and nucleic acids.

But the ribosome is complex, which is why the atomic-resolution crystal structure of the ribosomes in the bacterium *Escherichia coli* has been so long in coming. Jamie Doudna Cate, of the University of California, Berkeley, and his co-workers at last show us what the ribosomes of this archetypal prokaryote look like. The unprecedented level of detail gives a clearer picture of how the intricate protein factory functions.

SPL

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CANCER

Catch it quick

Proc. Natl Acad. Sci. USA **102**, 16368–16373 (2005)
Hopes of detecting cancers early and non-invasively by searching for mutant DNA in patients' blood plasma have been put to the test in a study of people with colorectal cancer.

The results, collected from 33 patients with colorectal tumours and 10 without, indicate that even low levels of mutant DNA from cancerous cells can be picked up. The technique identified malignant tumours, often before they spread, but not premalignant masses. The plasma DNA was analysed using a sequencing assay developed for the purpose, report researchers led by Bert Vogelstein of the Johns Hopkins Medical Institutions in Baltimore, Maryland. They propose that the abnormal DNA is released into the blood when tumour cells are engulfed by immune cells.

BIOTECHNOLOGY

Tiny tweezers

Lab Chip **5**, 1224–1228 (2005)

Living *Escherichia coli* bacteria can be teased into a three-dimensional arrangement inside a gelatin sample using optical tweezers, researchers have shown. The technique, invented by a team of scientists from Sweden and Britain, could be used to study cell function or to build templates for growing replacement tissue.

Tweezers made from infrared laser beams were used to move the bacteria around in the liquid gelatin before it set. Once the cells were fixed in place, the lasers were switched off. Provided with the appropriate nutrients, the bacteria could survive within the gelatin matrix for several days.

PHYSICS

Spins in sync

Nature Phys. **1**, 111–116 (2005)

Experiments have confirmed that the quantum characteristics of a Bose–Einstein condensate extend to the spin of its atoms.

Michael Chapman and colleagues from the Georgia Institute of Technology, Atlanta, made their quantum gas condensate by cooling rubidium atoms until they all dropped into the same quantum state. Using magnetic fields, the researchers were able to visualize the condensate's spin, which can point in one of three directions. They also saw that interactions between condensates with different spins were reversible, unlike spin interactions in a classical gas.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

M. THOMAS/SPL

CELL BIOLOGY

Tipped for protection

Genes Dev. doi:10.1101/gad1293805 (2005)

Telomeres, which cap the tips of chromosomes, inhibit the signalling triggered by double-stranded breaks in DNA, report Ted Weinert, of the University of Arizona in Tucson, and his colleagues. This explains how telomeres help to prevent the cell's DNA-repair machinery from fusing chromosome ends.

Weinert's group compared the response of yeast cells to a double-stranded break near to a telomere sequence with a break elsewhere. The telomere sequences seem to act as 'anti-checkpoints', suppressing the recruitment of the checkpoint proteins that normally suspend the cell cycle while the damage is repaired.

IMMUNOLOGY

The enemy within

Science **310**, 850–855 (2005)

A deeper understanding of how the immune system avoids attacking our own cells may point the way towards a therapy for the autoimmune disease multiple sclerosis.

The amino acid tryptophan has previously been associated with control of 'anti-self' immune reactions. Lawrence Steinman of Stanford University, California, and his colleagues found that its action is triggered by the enzyme indoleamine 2,3-dioxygenase. Cells carrying 'self' markers boost transcription of this enzyme, which breaks down tryptophan. In a mouse model of multiple sclerosis, the products of the tryptophan breakdown, and similar molecules, shut down anti-self reactions and eased symptoms of the disease.

JOURNAL CLUB

Steven Salzberg
University of Maryland,
College Park, USA

A genomics researcher wonders why viruses don't get the respect they deserve.

After sequencing hundreds of flu genomes this year (E. Ghedin *et al. Nature* **437**, 1162–1166; 2005), my colleagues and I constantly find ourselves discussing the flu virus as if it were a living organism. We try to be careful to say 'particle' instead of 'cell', and 'segment' instead of 'chromosome'. But are we right to imply that viruses are alive?

Like many students, I learned in high school that viruses were not alive. And I never had cause to question this until last year, when Jean-Michel Claverie visited my group to talk about the sequencing of the 'giant' virus known as mimivirus.

Claverie and his colleagues had just claimed in *Science* (D. Raoult *et al. Science* **306**, 1344–1350; 2004) that this organism has so many capabilities it deserves its own branch on the tree of life. The evidence is compelling: mimivirus has a double-stranded DNA genome of 1.2 million base pairs, with many genes never before seen in viruses, including several necessary for protein synthesis and even DNA repair.

Not surprisingly, this study provoked controversy. Some argued (D. Moreira & P. López-García *Science* **308**, 1114; 2005) that mimivirus is just an unusually adept 'gene pickpocket', scavenging genes from its hosts.

Claverie's team offered a spirited defence, saying the genes' ancient origin supports the idea that viruses have gradually trimmed down since they originated near the base of the tree of life, shedding genes to rely increasingly on their hosts.

Viruses have survived for millions of years, evolving and adapting just like 'real' life forms. But if we accept that they branched from the tree of life, we will face another dilemma: how much original genetic material must they have lost for us to stop calling them life? The tiny influenza virus, with just 11 genes, is waiting to see if we can figure it out.

NEWS

Deadly flu virus can be sent through the mail

The reconstructed version of the flu virus that caused the 1918 world pandemic will be mailed to registered labs that ask for it, despite previous assurances to the contrary, *Nature* has learned.

The DNA sequence of the virus, which killed some 50 million people when it swept the globe in 1918–19, was reported by US researchers last month (J. K. Taubenberger *et al. Nature* 437, 889–893; 2005). At the same time, another group described how it used that sequence to reconstruct the complete virus for the first time (T. M. Tumpey *et al. Science* 310, 77–80; 2005).

Critics such as virologist Jens Kuhn of Harvard Medical School say that the virus should never have been recreated in the first place, as it could escape and cause another pandemic. Giving the virus to more labs, not to mention sending it in the mail, adds to that risk, they say. Others argue that such studies are safe, and can help to discover what made the virus so deadly. “We have to be careful,” says US health secretary Michael Leavitt. “But on the other hand we have to have the ability to study it.”

Scientists at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, where the virus is held, initially answered concerns that it might escape by emphasizing that only one person, microbiologist Terrence Tumpey, had access to it (see *Nature* 437, 794–795; 2005). Tumpey said that he had undergone extensive background security checks, and had to pass through tough safety procedures every time he entered the lab. The CDC also said that the virus would not be sent out to any other labs.

The agency now seems to have changed its mind. Spokesman Von Roebuck told *Nature* last week that labs that are registered to work with select agents — in particular, dangerous pathogens that are subject to specific handling rules — will be able to request the virus. The reconstructed flu virus was added to the CDC’s select-agents list on 20 October.

The designation means that labs that operate under enhanced biosafety level-3 conditions or higher will be able to work with the virus. The highest biosafety level is 4, which requires full body suits. But enhanced level-3 conditions still require lab workers to wear respirators and to shower before leaving the lab. According to the select-agent rule, it is also

“strongly recommended” that lab workers are vaccinated with the annual flu vaccine, which may offer partial protection.

No lab has yet formally requested the virus, Roebuck says. But some scientists say they would like to work with it. Michael Katze, a virologist at the University of Washington in Seattle, says there are plans to renovate his institution’s primate centre so that his group can at least work with forms of the virus that contain some of the 1918 genes. Katze says he will study the virulence of the 1918 strain by infecting macaques with a mutated version of the virus. At some point, he also wants to work with the fully reconstructed virus, he says.

Scientists in Canada are planning to work with the virus, although they will not request it from the CDC. Constructs containing the virus’s DNA will be made at the University of Wisconsin-Madison and will be sent to the National Microbiology Laboratory in Winnipeg, Canada, according to Frank Plummer, the laboratory’s scientific director. Reconstructing the live virus from its DNA would then take just a few days, he says.

The researchers at the Winnipeg lab, which has level-4 certification, will infect mice and macaques to identify which parts of the virus make it so virulent. Such knowledge may enable researchers to identify potentially dangerous flu strains and develop vaccines against them. The work has been given extra urgency by the avian flu strains currently being carried by birds migrating from Asia. Researchers analysing the 1918 sequence have

concluded that the virus was one that had jumped to humans from birds.

If it is sent out, the 1918 virus could be shipped by commercial carrier services that allow it to be tracked, says Clarence Peters, a virologist at the University of Texas Medical Branch in Galveston. Frozen samples of select agents are shipped in a plastic vial wrapped in absorbent material that would soak up any

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

How bad would it be if the virus escaped?

Nobody really knows what the consequences would be if the reconstructed 1918 virus escaped or was released maliciously.

Some virologists point out that current flu drugs and vaccines are partially effective in mice against virus strains containing key genes from the 1918 flu. But Clarence Peters, a virologist at the University of Texas Medical Branch in Galveston, says the vaccine

has not been tested on the full virus, and points out that flu viruses could quickly mutate to become resistant to drugs given to researchers working with the 1918 strain.

There is little research on how the virus is transmitted between humans or on how long it survives on surfaces, adds Matthew Meselson, a biologist and bioweapons expert at Harvard University.

Andrew Pekosz, a virologist at

Washington University School of Medicine in St Louis, Missouri, says studies from the 1970s show that it can survive on surfaces for several hours. Peters also points out that if the 1918 flu virus infected someone, it would be much harder to contain than, say, Ebola, which requires close contact to pass from one person to another. “It would spread quickly, particularly given travel and transport today,” he says. **A.V.B.**



THE NATURE PODCAST
Listen to the stories behind
the hottest research in our
weekly show, available free
at [www.nature.com/
nature/podcast](http://www.nature.com/nature/podcast)

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

T. GLASGOW/AP

Making tracks: the 1918 flu virus could be transported between labs by commercial carriers.

leaks. That vial is placed inside at least one other plastic container. This is held inside a polystyrene box containing dry ice, which itself lies inside a heavy cardboard box. Some of the packaging must pass tests such as free-fall drops and has survived air crashes, says Plummer: "It's very, very safe."

Critics note that there is still a risk that the package could be lost or misdelivered, but Plummer says that is unlikely because the sender and recipient would know that the package is in transit. He adds that other select agents, including Ebola virus, are frequently sent in this way. If the virus did escape, it is unclear how serious the consequences would be (see 'How bad would it be if the virus escaped?').

Kuhn argues that the more places and people work with the virus, the greater the chance that it will escape or be stolen. He says that there should be an international agreement that restricts work on the virus to a few laboratories worldwide. ■

Andreas von Bubnoff
See Editorial, page 130

Y.-J. JUNG/AP/EMPICS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Far East lays plans to be stem-cell hotspot

Asia is aiming to secure its place as the world leader for cloning technologies. Many of the world's leading stem-cell biologists and cloning specialists hail from countries such as South Korea and Japan. And judging by the second Asian Reproductive Biotechnology conference, held from 2 to 7 November in Bangkok, Thailand, these pioneers are willing to share knowledge and techniques with scientists from less developed neighbours in the region, who are keen to enter the game.

"There is huge potential for Asian scientists here," says Woo Suk Hwang of Seoul National University in South Korea, who last year led the first team to derive stem cells from a cloned human embryo.

Strong government support and relatively relaxed ethical regulations are often cited as reasons for Asia's success in stem-cell biology, but Hwang singles out technical prowess as a critical factor. Many stem-cell biologists have noted that some labs in Europe and the

United States have struggled with skills already mastered by researchers in Asia, such as those needed to operate the micromanipulator devices used to remove nuclei from cells during the cloning process.

The researchers at the meeting, who numbered around 150, saw a demonstration of one such technique from Teruhiko Wakayama of the RIKEN Center for Developmental Biology in Kobe, Japan, who in 1998 created the first cloned mouse. Biologists from Vietnam and Thailand were able to try extracting nuclei from egg cells with his piezoelectric device.

Many labs in Asia, in which even basics such as refrigerators can run short, cannot afford to experiment with such techniques, says Nguyen Van Thuan, who also works at the RIKEN Center. He presented one potential solution: a method for storing mouse sperm at room temperature for up to a week by adding salt to the bovine serum albumin solution in

which the sperm is often held (N. Van Thuan *et al. Biol. Reprod.* 72, 444–450; 2005). Thuan, who is Vietnamese, says that the technique could help his country's scientists to transport samples and run studies involving mice sperm, without taking up scarce refrigerator space.

Despite a lack of resources, Thailand and Vietnam — host country for the first Asian Reproductive Biotechnology meeting — are hoping for a bright future. The Thailand Research Fund will start a US\$50-million grant programme for stem-cell research in 2006, and earlier this year the Vietnamese government promised to boost research spending to 2% of its gross domestic product (GDP). This will more than double current investment, and a good chunk is expected to go to reproductive biology. Two stem-cell scientists, Nguyen Mong Hung at the Hanoi University of Science, and Bui Xuan Nguyen at the Vietnamese Academy of Science and Technology in Hanoi, will be getting US\$500,000 a year to head centres of excellence in the field.

Hwang says that there is particular scope for collaboration in the cloning and derivation of monkey stem cells to develop models for human disease. He expects that a collaboration of scientists from South Korea, Japan, China and other Asian countries with access to primate populations will form a significant part of his recently formed network for exchanging stem-cell lines and cloning technology (see *Nature* 437, 1077; 2005). ■
David Cyranoski

Woo Suk Hwang says Asia has huge potential in cloning.

ON THE RECORD

“We ran aground on a coral reef we were trying to protect.”

Greenpeace says sorry after its boat, *Rainbow Warrior II*, hits an ecologically fragile reef in the Philippines.

“I find it inconceivable that this paper is not well known.”

Cornell physicist Neil Ashcroft is surprised to discover an obscure 1922 paper on superconductivity — by Albert Einstein. A translation of the paper is now on the arXiv physics preprint server.

Sources: CNSNews.com, PhysicsWeb

SCORECARD**Memorials**

An art company based in Japan is offering a fresh twist for gardens of remembrance. It plans to make ‘living tombstones’ by generating trees whose every cell contains the DNA from a deceased loved one.

**Child prodigies**

An eight-year-old boy who dreams of building flying cars and joining the European particle-physics laboratory, CERN, has become the youngest pupil to enrol at a South Korean university.

**Jet lag**

Researchers in Chicago believe they have come up with a potent method for resetting travellers’ body clocks. They say that a combination of bright light and melatonin has a much stronger effect than either element on its own.

NUMBER CRUNCH

A survey by the American Association for the Advancement of Science reports that 40% of its members have had ‘difficulties’ acquiring patented technologies to use in their work. Among those who had problems:

58% said that their work was delayed by the difficulties.

50% said that the problem forced them to change their research.

28% had to abandon their project altogether.

Source: <http://sippi.aaas.org/survey>

L. PATERSON/SPL

Regulations on experiments involving human subjects are not always followed to the letter.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Researchers break the rules in frustration at review boards

The watchdogs that oversee the ethics of human research projects can sometimes provoke scientific misconduct. That is the counter-intuitive conclusion of a series of papers to be published over the next few months. The authors, who specialize in research ethics, say they have evidence that some ethics panels are alienating researchers and inadvertently promoting deceit.

Patricia Keith-Spiegel of Simmons College in Boston, Massachusetts, says she began her studies after hearing of cases at other US institutions where scientists had violated research rules after feeling that they had been mistreated by institutional review boards (IRBs). Experiments involving human subjects in the United States, from social-science studies to medical research, must be rubber-stamped by an IRB. Researchers acknowledge that the boards are

necessary to ensure that subjects are treated correctly, but sometimes complain that the boards fail to understand the research involved and do not explain their decisions properly.

As an example, Keith-Spiegel cites a researcher she knows who became frustrated at lengthy IRB review times and so routinely

began data collection before receiving approval. Another researcher admitted to omitting aspects of protocols for research projects after receiving demands for numerous “picky” changes. Typical IRB requests include changes to

consent forms or restrictions on the type of questions that subjects can be asked.

“I realized that there are scientists who want to do things the right way but who are having to distort their research protocols because of perceived unreasonable or ridiculous demands from IRBs,” says Keith-Spiegel.

“Researchers are more open to committing misconduct if they feel wronged by a review board.”

These fears are backed up by a survey of misconduct rates among 3,000 researchers funded by the US National Institutes of Health. Published earlier this year, it found that a third of respondents had engaged in one of ten types of misconduct in the past three years (see *Nature* **435**, 718–719; 2005, and B. C. Martinson *et al.* *Nature* **435**, 737–738; 2005). Further analysis of the survey data, to be published next March in the *Journal of Empirical Research on Human Research Ethics*, shows that misconduct rates were highest among researchers who felt that they had been unfairly treated by other governing bodies in science, such as funding review panels. A similar relationship is likely to exist between misconduct and the perception of unfair treatment by IRBs, says Brian Martinson of the HealthPartners Research Foundation in Minneapolis, Minnesota, lead author on the two studies.

Keith-Spiegel has also studied the issue by asking scientists' opinions on fictional situations in which an IRB refused researchers permission for a study. In cases where the IRB responded in a curt manner, rather than explaining its decision, subjects empathized with the rejected researcher and assigned a less significant punishment if that researcher went ahead and ran the study anyway. The results are still being analysed, but Keith-Spiegel says they seem to suggest that researchers are more open to committing misconduct if they feel wronged by an IRB.

Keith-Spiegel and Martinson say that their findings can be explained by organizational justice theory, a well-established method for studying workplace relationships. Studies in work environments other than science have shown that employees are more likely to commit misconduct if they feel their managers are not giving them due reward or are treating them unfairly. In a paper due to appear in next month's *Ethics and Behavior*, Keith-Spiegel argues that the same relationship can exist between IRBs and scientists.

Ethics committee chairs who spoke to *Nature* say they try to avoid such problems by maintaining a good relationship with scientists. "I've certainly heard of problems," says Leigh Firn, chairman of an IRB at the Massachusetts Institute of Technology. "But we don't see ourselves as the ethics police. Unless it is something of substance we won't request changes." Brian Shine, a consultant at Oxford Radcliffe Hospitals Trust, UK, and chairman of a local ethics committee, adds that he invites researchers to meetings to discuss potential problems and always writes to them afterwards to clarify the discussion. ■

Jim Giles

Boeing strike leaves satellites stranded on launch pad

A machinists' strike is hitting some US space projects hard. It has already delayed the launch of three atmospheric satellites and it could derail a major Pluto mission if it is left unresolved.

About 1,500 Boeing machinists and engineers walked off the job at facilities across the United States on 2 November, after talks between the union and the company on health care broke down. The machinists are responsible for the assembly and launch preparation of the company's Delta rockets, commonly used in scientific missions for NASA.

The strike stopped the countdown of the National Oceanic and Atmospheric Administration's hurricane-tracking satellite GOES-N, scheduled to launch on 5 November, leaving it stranded atop its Delta IV rocket in Cape Canaveral, Florida. Boeing officials are now assessing whether it will be possible to restart the countdown using non-union employees, says Robert Villanueva, a spokesman for the company.

And delayed indefinitely are NASA's CloudSat and CALIPSO satellites, which will study the global distribution of aerosols (see *Nature* **437**, 468–469; 2005). They were set to launch on a single Delta II rocket in November, but those plans are now on hold. David Winker, principal investigator for CALIPSO, says that the delay is especially worrying because the satellites are meant to take part in international projects in which many teams collect climate data at the same time. "These things are going to go ahead whether we launch or not," he says. "It's close to being critical."

If the strike becomes protracted, it may even affect a mission to Pluto. New Horizons would be the first spacecraft to visit the Solar System's most distant planet, and the final stage of the vehicle is propelled by a Boeing engine. If the mission misses its month-long launch window early next year, its next chance will not be until February 2007. But Villanueva says it should be possible to complete the necessary work using replacement technicians and inspectors.

He adds that there is no schedule for resolving the strike: "Basically both sides are sitting back and leaving lines of communication open." ■

Geoff Brumfiel



PROBIOTICS GET A BOOST

Sellers of 'friendly' bacteria offer evidence to back health claims.

www.nature.com/news

Antigravity craft slips past patent officers

The US patent office has granted a patent on a design for an antigravity device — breaking its own resolution to reject inventions that clearly defy the laws of physics.

This is not the first such patent to be granted, but it shows that patent examiners are being duped by false science, says physicist Robert Park, watchdog of junk science at the American Physical Society in Washington DC. Park tracks US patents on impossible inventions. "The patent office is in deep trouble," he says.

"If something doesn't work, it is rejected," insists Alan Cohan, an adviser at the patent office's Inventors Assistance Center in Alexandria, Virginia. And when something does slip through, he says, the consequences are not significant: "It doesn't cause any problems because the patent is useless."

But Park argues that patenting devices that so blatantly go against scientific understanding could give them undeserved respectability, and undermine the patent office's reputation. "When a patent is awarded for an idea that doesn't work, the door is opened for sham."

Patent 6,960,975 was granted on 1 November to Boris Volfson of Huntington, Indiana. It describes a space vehicle propelled by a superconducting shield, which alters the curvature of space-time outside the craft in a way that counteracts gravity. The device builds on a claim by the Russian physicist Eugene

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Balls up: US patent office falls for antigravity device that would allow perpetual-motion machines.

Podkletnov that superconductors can shield the effects of gravity. NASA was at one stage investigating the idea, but it has become almost as notorious as cold fusion as an example of fringe science.

One of the main theoretical arguments against antigravity is that it implies the availability of unlimited energy. "If you design an antigravity machine, you've got a perpetual-motion machine," says Park. Shield half of a wheel from gravity and it will keep turning for ever.

The US patent office has long fought to prevent applications for patents on perpetual-motion machines. In 1911, after a constant stream of applications, one commissioner ruled that they would not be considered until a working model had been running for a year. More recently, inventor Joe Newman sued the office after it rejected his application for such a device. The court finally ruled against Newman in 1990, a decision that the patent office cites in its rules about which inventions are patentable.

Unfortunately, it is not always easy to tell what the implications of a patent are. One previous patent for a device using putative "hydrinos" — shrunk hydrogen atoms — to produce huge amounts of energy was granted. It is currently being reviewed after several scientists complained that hydrinos are impossible according to the laws of physics.

Park says he sympathizes with the difficulties that patent examiners face. "Their burden has gone up enormously," he says. "It's not surprising they get in a jam."

Philip Ball

Bush buries US bunker-buster project

WASHINGTON DC

Bowing to congressional pressure, the administration of President George W. Bush has killed a programme to design a nuclear warhead capable of striking targets buried deep in the ground. It has instead chosen to develop replacement warheads for the existing nuclear stockpile.

The cancelled weapon, called the Robust Nuclear Earth Penetrator (RNEP), was proposed three years ago (see *Nature* 415, 945; 2002). It was to have been a toughened version of an existing warhead that could strike bunkers and other underground targets.

However, some physicists were sceptical that the warhead, known

as the 'bunker-buster', could penetrate deeply enough to contain its massive nuclear blast. A National Research Council study released in May showed that the weapon would be highly effective at destroying deeply buried targets. But casualties could still number in the thousands, says John Ahearne of the scientific research society Sigma Xi, who chaired the study.

Within Congress, Republicans and Democrats alike opposed RNEP development on both political and technical grounds. Last year, a bipartisan coalition killed funding for the programme. In the face of such opposition, the administration has essentially

pulled the plug on the project by withdrawing the \$4 million requested to study RNEP, according to Senator Pete Domenici (Republican, New Mexico).

"The focus will now be with the defense department and its research into earth-penetrating technology using conventional weaponry," Domenici said in a statement.

"This decision simply confirms what the critics have been saying all along," says Christopher Paine, a senior analyst at the Natural Resources Defense Council. "The programme had become a thorn in the side of the administration."

But as RNEP disappears, another project is on the rise.

A congressional budget bill expected out this week would allocate \$25 million to speed up the Reliable Replacement Warhead project. The programme, established to design the next generation of weapons for the ageing nuclear stockpile, has itself been generating controversy among researchers, some of whom believe it to be unnecessary (see *Nature* 434, 684; 2005).

"RNEP may be politically dead," says Daryl Kimball, executive director of the Arms Control Association in Washington DC. "But I would be very surprised if the debate over new weapons ends any time soon."

Geoff Brumfiel

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Under threat: the law that has helped protect the grizzly bear for more than 30 years may be scaled back.

Congress attacked over species bill

SAN DIEGO

Conservationists say members of Congress are misrepresenting science in a bid to change the way endangered species are protected in the United States.

The record of the Endangered Species Act (ESA) is being twisted to win passage of a new law that would reduce protections for animals and plants, leading biologists say. They are now calling on other scientists to get involved in the debate to try to defeat the measure, which has passed the House of Representatives and is soon to emerge in the Senate.

Since it was passed in 1973, the ESA has been the country's most important law for designating species as threatened or endangered, and ensuring that landowners and industry minimize damage to habitats while logging, mining or developing land.

Many industry groups, with support mainly from Republicans, are keen to scale back environmental protection laws. One particular supporter is Representative Richard Pombo (Republican, California). Pombo chairs the House Committee on Resources, and last May his staff released a report arguing that the ESA is failing to protect endangered species. The report points out that only about a dozen of the more than 1,300 species protected by the law have made it off the threatened and

endangered lists, concluding that the law "does not appear sustainable".

In September, Pombo introduced a bill to the House that suggested major changes to the act. These included weakening the requirements for recovery plans for species, permitting the secretary of the Department of the Interior to overrule any scientific decision on how to save a species, and allowing the elimination of habitat protection shown to be vital for species recovery. The bill passed on 21 September with little discussion.

Rule benders

But experts in the field have complained that scientific results were consistently ignored or misrepresented by Pombo's team. "The goal of this legislation is to emasculate the ESA," says conservation biologist Dennis Murphy of the University of Nevada, Reno. "The legislation has nothing to do with science, and everything to do with economics."

"The resources committee report is a biased, unbalanced representation of the ESA," agrees wildlife biologist Barry Noon of Colorado State University in Fort Collins, who studies threatened species. Noon says it is misleading to call the law a failure just because only a few species have made it off the threatened and endangered lists.

By the report's rationale, says Noon, many of the nation's most famous conservation successes are failures, including the dramatic resurgences of the bald eagle (*Haliaeetus leucocephalus*) and the grizzly bear (*Ursus arctos*), both of which are still listed.

Noon points out that a peer-reviewed report by the lobby group Environmental Defense shows that the conservation status of 52% of listed species is improving (T. D. Male and M. J. Bean *Ecol. Lett.* 8, 986–992; 2005). And the evidence available for the 40% of species that haven't been fully surveyed because of lack of funds suggests that the status of many of these is improving too.

The populations of virtually all the listed species were small and in decline when they were put on the list, Noon notes. It can be very difficult to halt declines quickly, he adds, because it takes time to change or eliminate agricultural and industrial practices. And the biology of certain species can mean that even under the best conditions, their recovery can take decades if not centuries.

"One needs to look at the recovery rate relative to the life history of a species," he says. "That includes the rate of reproduction, the age that needs to be reached for reproduction, the size of the population and the nature of the threat to the species."



WARNING SHOT FOR GREEN CHEMISTRY

Some solvents with an environmentally friendly reputation may kill fish.

www.nature.com/news

As examples, environmental groups have pointed to the recovery of the Northern right whale (*Eubalaena glacialis*) and the red-cockaded woodpecker (*Picoides borealis*). Government officials, scientists and industry representatives estimate that it will take 150 years to recover the right whale, nearly extinct after 1,000 years of whaling, and 70 years to rescue the red-cockaded woodpecker, endangered by loss of its forest habitat from 300 years of logging.

Scientists at the Center for Biological Diversity in Tucson, Arizona, an environmental organization that has used lawsuits to prompt enforcement of the ESA, say their analysis of government data shows that the average expected recovery time for species currently on the list is 35 years.

In June — shortly after the Resources Committee report was published — Murphy, Noon and botanist Bruce Pavlik of Mills College in Oakland, California, travelled to Washington DC to explain their concerns to Pombo's staff. But Noon says that the staff

paid little heed to their concerns.

"I was mad," he recalls. They weren't interested in what we had to say." Pombo's aides declined *Nature's* request for an interview.

Pombo, who did not meet the three scientists, defends his bill, saying that all the material in the committee's report came from government agencies whose responsibilities are to monitor species. "We are open to everybody and anybody," he says, adding that "dozens and dozens of biologists have testified" at committee hearings on the ESA during the past 12 years.

Murphy, who sometimes works with industry to find productive solutions to conservation issues, admits that the ESA has shortcomings. "Substantial changes need to be made," he says. "But those types of change aren't in this bill."

First, the federal agencies currently responsible for determining whether a species should be listed "don't have the technical expertise to make the decision," he says. The habitat that species need to recover should also be defined more rigorously, he adds. "We often don't

identify the specific resource needs for a species to survive in the long haul." Improving this could reduce litigation over protecting species on private lands: "We like to think that better science on this will get us a better ESA."

A companion bill to Pombo's is expected to be introduced to the Senate later this month, and will initially be dealt with by an environmental subcommittee chaired by Lincoln Chafee, a moderate Republican from Rhode Island. Debate in the Senate may be influenced by further discussions due to be held on the ESA, including meetings planned for November and December, which will be hosted by the Keystone Symposia in Colorado. And the US Government Accountability Office, a federal agency that conducts politically neutral reviews, is expected to complete by March a report on the process the ESA uses to come up with recovery plans.

In the meantime, Noon hopes that more scientists will get involved in the debate. "Scientific information is power," he says. ■

Rex Dalton

With additional reporting from Emma Marris, Washington DC

"One needs to look at the recovery rate relative to the life history of a species."

US cheered by rise in foreign student numbers

For the first time in several years, the number of foreign graduate students enrolling at US universities has increased.

Strict visa regulations, put in place after the 11 September 2001 terrorist attacks, have caused the number of international students to drop dramatically in recent years (see *Nature* 431, 231; 2004). But first-time enrolments were up 1% in the 2004–05 college year, says a survey by the non-profit Council of Graduate Schools, based in Washington DC. The most dramatic turnaround was for Chinese students, who make up the largest group. They rose by 3%, after an 8% drop the previous year.

The survey of more than 125 institutions contained mixed results for the sciences. Physical sciences received a 1% boost in enrolment, whereas life sciences saw their enrolments drop by 1%. But that was better than last year, when first-time enrolments in life-sciences programmes had plummeted by 10%.

“This is a very positive sign,” says Heath Brown, director of research and policy analysis for the council. “But we’re still a considerable way from where we’d like to be.”

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hanan Eshel bought a fragment of an ancient Torah scroll from some Bedouin Arabs.

Dead Sea scroll lands Torah expert in hot water

A respected Israeli archaeologist is being investigated by police over his handling of an ancient scroll.

Last February, Hanan Eshel of Bar-Ilan University near Tel Aviv bought a fragment of an ancient Torah scroll from three Bedouin Arabs. The Arabs said they found it in a cave in the Judean Desert — the same area where the Dead Sea scrolls were found. Based on the style of lettering and on finds later uncovered in the same cave, Eshel dated the scroll to about 135 AD.

Under Israeli law, all archaeological finds belong to the Israel Antiquities Authority. The agency says Eshel should have turned over the find within two weeks of receiving it — not months later, as he did. But Eshel argues that he should be commended for rescuing an artefact that was in danger of decay or of being sold to a collector overseas. “In terms of scholarship, this is a success story,” he says.

Hunt for uncle hopes to confirm Copernicus find

Polish archaeologists have launched a search for the relatives of Nicolaus Copernicus. They hope to run a DNA check on recently unearthed remains thought to be those of the sixteenth-century astronomer.

The researchers excavated a skull and bones last year from a grave in the Gothic Cathedral of Frombork on Poland’s Baltic coast, where Copernicus served in the clergy. On 3 November forensic experts in Warsaw released a computer-generated facial reconstruction. The face of the 70-year-old man bears a striking resemblance to Renaissance portraits of Copernicus when he was about 40.

But only genetic comparison with a close

M. KAHANA/AFP/GETTY IMAGES

relative will confirm the identity of the bones, says Jerzy Gassowski of the Institute of Anthropology and Archaeology in Pultusk, who led the excavation. Copernicus had no children so the scientists are focusing on his maternal uncle, Lukas Watzenrode, who died in Torun in 1512 and is thought to be buried nearby.

Heavyweight panel needed to protect biodiversity

Scientists are calling for the creation of an international, interdisciplinary panel to build global support for biodiversity.

Researchers are meeting this week at a conference in Oaxaca, Mexico, that is sponsored by DIVERSITAS, an organization that aims to ensure that biodiversity is central to policy decisions. The group has been discussing the setting up of a body modelled on the Intergovernmental Panel on Climate Change (IPCC), which assesses and issues reports on climate science.

"The IPCC has moved climate change higher up the international political agenda, because it provides timely independent scientific advice on which to base policy decisions," says Anne Larigauderie, executive director of DIVERSITAS. "This is

Confusing signal keeps Japanese craft hanging around

Japan's space agency last week cancelled a practice touchdown on the asteroid Itokawa. The Hayabusa spacecraft is currently orbiting Itokawa, and had been scheduled to descend to the surface on 12 November for the first of two visits to collect samples.

But on 4 November, an anomalous communication



caused mission controllers to stop the practice descent. As *Nature* went to press, the agency had not announced how the glitch might affect the scheduled touchdowns. In the meantime, Hayabusa continues to send back detailed photos of the 540-metre-long Itokawa, making it one of the best-studied asteroids yet.

ISAS/JAXA

what the biodiversity community would like." The idea for an international biodiversity panel was first suggested at a meeting in Paris in January.

Chinese observatory gives a glimpse of the past

Chinese archaeologists have unearthed an observatory thought to be 4,100 years old. The remains, close to the city of Linfen in Shanxi province, include a semicircular platform about 40 metres across surrounded by 13 pillars.

"People observed the direction of the sunrise through the gaps between the pillars, and distinguished the different seasons of the year," archaeologist He Nu of the Chinese Academy of Social Sciences told the Xinhua news agency.

After 18 months of study, the researchers concluded that the periods represented by the 12 gaps between the pillars matched the way the seasons are divided in the traditional Chinese calendar. Accounting for changes in Earth's axis of rotation, the scientists determined that the midwinter sunrise would have occurred in the middle of one of the gaps 4,026 years ago.

More than a cosmetic change

Commercial and political pressures are pushing for a halt to the use of animals in toxicology tests in Europe. This change will also mean a move towards better science, says **Alison Abbott**.

Every time you reach for an eyedrop or reapply a lip salve, you do so confident that the chemicals they contain are safe to use. But the toxicology tests on which regulators rely to gather this information are stuck in a time warp, and are largely based on wasteful and often poorly predictive animal experiments. Efforts in Europe are about to change this, and the man charged with bringing toxicology into the twenty-first century is a plain-talking German: Thomas Hartung. Although Hartung acknowledges the immense challenges ahead, he sees this as an opportunity for toxicology "to turn itself at last into a respectable science".

Three years ago, when Hartung became director of the European Centre for the Validation of Alternative Methods (ECVAM) in Ispra, Italy, he didn't know that the job was about to shift gear dramatically. ECVAM was set up in 1993 to support European Union policy aimed at reducing the number of animals used in regulatory testing.

The centre, which nestles on the sleepy shores of Lake Maggiore in the Italian Alps, originally had ten members of staff and faced an uphill struggle to cut back the millions of animal tests carried out in Europe every year. Then in 2003, two major policy changes were announced from above, increasing the pressure on the centre's labs. ECVAM found itself facing an unexpectedly short deadline for delivering a slew of animal-free methods for testing chemical toxicity.

Rule change

The first change was to the European Union's Cosmetics Directive, which phases out over ten years the use of animals in cosmetics testing. A short while later, the European Commission proposed its controversial REACH legislation (Registration, Evaluation and Authorization of Chemicals). Europe produces some 30,000 chemicals for which toxicity data have never been registered. REACH aims to make registration mandatory for both future and existing chemicals — even those that have been on the market for decades.

If, as expected, the REACH directive is approved next year, it will come into effect in 2007. Animal-welfare groups fear that this will mean millions more animals will be used in tests to meet the regulatory requirements. And industry claims that the testing process could cost it billions of euros. Almost overnight, industry's interest in cheaper, animal-free testing skyrocketed.

Last month ECVAM was put in charge of



Skin deep: animal testing for all cosmetics is being phased out in Europe.

developing, with industry and regulatory agencies, the testing strategies for REACH. Now commanding 50 staff, Hartung is rising to the challenge. The toxicity tests that have been used for decades are "simply bad science", he explains. "We now have an opportunity to start with a clean slate and develop evidence-based tests that have true predictive value."

Many of the animal tests used today were developed under crisis conditions. The notorious Draize test, which assesses the irritation or damage caused by chemicals simply by putting them into the eyes of rabbits, is a prime example. It was developed by the US Food and Drug Administration in 1944 after reports in the 1930s that some cosmetics were causing permanent eye injuries. One 38-year-old woman had gone blind after dyeing her lashes with Lash-Lure, a product that contained a derivative of coal tar.

Then came the calamity of thalidomide, which was given to pregnant women in the late 1950s to control morning sickness, but which caused horrific birth defects. By this time, gov-

ernments were highly sensitive to public concerns and called on their authorities to develop animal-based tests that would predict all conceivable toxic effects of drugs and chemicals. The principles behind most of those tests remain more or less unchanged today.

The battery of tests demanded by European authorities covers all eventualities, from acute effects that are seen shortly after exposure — such as eye and skin irritation — to concerns about whether in the longer term a compound might cause cancer, or brain or birth defects. In addition, tests must be carried out to define the risk a chemical might pose to the environment, such as toxicity to fish and other aquatic species.

Safety catch

Each chemical that goes through the multiple tests required for registration can use up to 5,000 animals — or 12,000 if the chemical is a pesticide. The cost of doing this for the 30,000 unregistered chemicals so that they comply with REACH has been estimated at between €5 billion (US\$6 billion) and €10 billion.

In the decade since ECVAM was established, the number of animals used in toxicology testing has fallen slightly, although it still hovers at about one million per year. This reduction is a result of the refinement of existing tests, and the introduction of some alternative methods that rely on *in vitro* tests using cell cultures.

ECVAM is still pushing on both fronts. For

"To test a chemical for its potential to cause cancer takes five years and involves 400 rats. More than 50% of the results are positive, of which 90% are false positives."

SHERMAN/GETTY IMAGES

example, the LD₅₀ acute toxicity test, which involves feeding animals with a chemical to determine the lethal dose, still accounts for one-third of all animal tests worldwide. But the numbers involved have fallen from 150 animals per chemical in the 1970s to just eight animals in 2002.

ECVAM believes that it can halve the total number of animals used for regulatory testing within a decade. It has just completed its first large-scale validation study of an *in vitro* cytotoxicity test, which monitors death of cultured cells following short-term exposure to a chemical. Chemicals shown to be harmful in this test would be excluded from any LD₅₀ animal tests. At least 70% of the chemicals registered in the past two decades fall into this category, says Hartung. And this is just the beginning.

"We've made a lot of progress with the low-hanging fruits, but many of the remaining animal tests — particularly the tests for toxicity in the long term — are much more challenging to replace," says Hartung.

Poor prediction

This is despite the acknowledged poor quality of most animal tests, which have never undergone the rigours of validation that *in vitro* alternatives now face. Most animal tests over- or underestimate toxicity, or simply don't mirror toxicity in humans very well.

Take the embryotoxicity test in which chemicals are fed to pregnant animals and the fates of their embryos, and the progeny of two subsequent generations, are studied. "Animal embryotoxicity tests are not reliably predictive for humans," says Horst Spielmann, a toxicologist at the Federal Institute for Risk Assessment in Berlin. "When we find that cortisone is embryotoxic in all species tested except human, what are we supposed to make of them?"

The same goes for cancer. To test a single chemical for its potential to cause cancer takes

Tests that put chemicals into the eyes of rabbits have changed little since the 1940s.

five years and involves 400 rats, each of which is treated with the maximum tolerated dose. It is dramatically over-predictive: more than 50% of the results are positive, of which 90% are false positives¹. Yet the number of compounds proved to be carcinogenic to humans is very low — the International Agency for Research on Cancer in Lyons, France, has identified just 95 proven and 66 probable human carcinogens.

As their experience grows, ECVAM staff are finding out just how complicated it is to develop persuasive alternative tests. A highly strategic approach is required, says Hartung. "In most cases it is not just a question of replacing one animal test with one *in vitro* test," he explains. Instead, a number of tests will be required, each of which has been

shown to match data on toxicity in humans, assuming such information is available.

For example, several *in vitro* alternatives have been developed to replace the Draize test, each with its own advantages and limitations. Among these, some are better at predicting mild irritation than physical damage. Others are more suited to a particular chemical class, such as detergents.

Life or death

Scientists also cannot assume that *in vitro* alternatives are automatically better, says Spielmann. In 1971, a comparison of animal Draize tests in different labs revealed the test to be hopelessly non-reproducible². But Spielmann's 1995 study of animal-free alternatives to the Draize test showed that they were equally unreliable³. Since then the *in vitro* tests have been standardized, and they are intrinsically more reproducible. "Although reproducibility and relevance are not the same thing," Spielmann cautions.

Relevance requires a good match between the test results and human data. At an ECVAM workshop in February, 30 industrial scientists met to develop the most effective strategy for using the alternative Draize tests, so that the false negatives and false positives of each test compensate for each other. This strategy is now going through the crucial validation procedure, in which human data, often from occupational health databases, will be used as points of reference (see 'The validation game', overleaf).

ECVAM has so far seen 17 alternative tests through validation — 11 use *in vitro* methods, another six involve refining *in vivo* tests to reduce the number of animals used. An additional 40 or so tests are under peer review, with more to come.

Most of the new tests assess acute toxicity,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CREDIT

ECVAM



Life-savers: researchers at ECVAM are developing alternatives to animal toxicity tests.

The validation game

The number of toxicity tests required by regulatory authorities mushroomed in the 1960s and 1970s. But tests accepted by authorities in one country were rarely identical to those accepted elsewhere. This forced companies to repeat the same test with small variations in every country they wanted to market their product, wasting both money and animals.

So in the early 1980s, the Organisation for Economic Co-operation and Development (OECD) launched a major harmonization initiative. National authorities met regularly under OECD auspices to negotiate modifications to their test schedules that would make them compatible with those in other countries. Most of the differences have now been resolved.

But at the same time there was growing pressure from the public to replace or reduce animal testing, so the OECD expanded its role to oversee global acceptance of new animal-free, or animal-lite, tests. Validation — the proof that a test accurately predicts a specific effect in humans — is the biggest

challenge for alternative methods.

At the European Centre for the Validation of Alternative Methods (ECVAM) in Ispra, Italy, validation typically requires testing a new protocol in three or four different external laboratories. The test chemicals are analysed by personnel who do not know the compounds' identities. If a test proves reproducible, it is sent for peer review by ECVAM's Scientific Advisory Committee, whose members include scientists, representatives from all member states of the European Union and relevant industrial and animal-welfare groups.

Once approved by the committee, the test is adopted by the European Chemicals Bureau, also based at Ispra, and then sent to the OECD for the all-important global validation. Most of the labs involved in ECVAM testing are European, but the centre has been careful to involve other regulatory authorities in early stages of the development process to speed OECD validation.

In 2000, the 3T3 Neutral Red



Uptake Phototoxicity Test became the first replacement test to be validated by European authorities.

The test identifies whether a chemical becomes toxic when exposed to light. Chemicals are added to a culture of skin cells, which is then irradiated with ultraviolet light. The rate at which the cells die before and after irradiation is monitored. The test was developed by Horst Spielmann from the Federal Institute for Risk Assessment in Berlin.

Ideally, the predictive value of a

new protocol is tested against human rather than animal data.

Human toxicity data are, of course, hard to find, so Spielmann was gratified to be invited to join in a clinical trial assessing the potential phototoxicity of drugs to which physicians had raised concerns. He was able to compare his test's data for some standard chemicals with the irritation they caused on a patch of human skin. He found a perfect match. The test was accepted by the OECD last year.

One of the 40 or so tests now going through validation is the new cytotoxicity test to help replace the animal lethal-dose (LD₅₀) test (see main story). It was the first validation study to involve both US and European groups from the start. It is also the first to use data from the records at national poison centres. The predictions of the *in vitro* test provided a better match than the rat LD₅₀ test when compared with the toxicity information on 42 chemicals listed as having poisoned people.

ECVAM

but animal use is highest when testing for the toxic effects of prolonged exposure to chemicals for long-term consequences such as cancer and reproductive toxicity. These costly procedures are harder to mimic *in vitro* and may never be completely replaced.

Sounds familial

This is why, apart from the €30 million it uses to support ECVAM annually, the European Commission is funding three multimillion-euro 'Integrated Projects'. Under these, dozens of labs will collaborate for five years to tackle more difficult issues, such as allergic reactions or widespread toxicity resulting from chemicals entering the bloodstream.

Scientists know that they are likely to find it hardest to convince regulators about alternative tests for highly emotive issues such as cancer and birth defects. More than half of all animals that will be needed to support REACH legislation are likely to be used in reproductive toxicology testing. The €9-million Integrated Project called ReProTect has 27 labs dedicated to developing alternatives to these tests. The ReProTect consortium has broken down the human reproductive cycle into smaller elements, from male and female fertility to implantation, to pre- and postnatal development, and is trying to develop a meaningful package of tests⁴.

"Quite correctly everyone feels uneasy about

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Rats are widely used to assess whether compounds can cause cancer.

taking risks where stakes are so high and issues so emotive," says Hartung. "We all want to be sure that there is real evidence that alternative tests are predictive of human toxicity."

For example, regulators know the weaknesses of the rat cancer test as well as scientists



"We all want to be sure that there is real evidence that alternative tests are predictive of human toxicity."
— Thomas Hartung

but, wanting to be safe rather than sorry, they accept it because it is believed to throw up few false negatives. They prefer to let industry prove the innocence of any compound that shows up positive. Any replacement tests will need to reassure both regulators and industry.

With so much yet to be achieved, is Europe on target to deliver toxicity tests to meet the new regulations? The Cosmetics Directive phases out animal use in acute toxicity testing in 2009, and in testing for long-term effects in 2013. "We are on target for the first deadline, but the second deadline may be more difficult," says Hartung. Fortunately, the 2013 deadline can be renegotiated.

The REACH legislation is yet to be finalized, and alternatives to tests that present the highest financial and animal burden, such as reproductive toxicology and carcinogenicity, will not be in place when REACH first becomes law. But the longer-term picture should see a reduction in animal suffering going hand in hand with use of better science.

Alison Abbott is Nature's senior European correspondent.

1. Gold, L. S., Manley, N. B., Slone, T. H., Rohrbach, L. & Garfinkel, G. B. *Toxicol. Sci.* **85**, 747–808 (2005).
2. Weil, C. S. & Scala, R. A. *Toxicol. Appl. Pharmacol.* **19**, 276–360 (1971).
3. Balls, M., Botham, P. A., Bruner, L. H. & Spielmann, H. *Toxicol. In Vitro* **9**, 871–929 (1995).
4. Hareng, L., Pellizzer, C., Brenner, S., Schwarz, M. & Hartung, T. *Reprod. Toxicol.* **20**, 441–452 (2005).

P. GOETHEL/UCK/SPL

ECVAM

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

TONGUE TIED

Endangered languages often contain key linguistic insights found nowhere else. But the tongues are disappearing faster than scientists can document them. **Jessica Ebert** reports.

The tips of dried sage burn in a small cast-iron pan like a row of lit matches. Alex Gwin, an elder of the Hidatsa Native American tribe, carries them from room to room chanting softly in his native tongue. The tiny two-bedroom house on North Dakota's Fort Berthold Reservation fills quickly with the sweet smoke.

Gwin extinguishes the sage twigs and sits at the dining-room table across from John Boyle, a linguist at the University of Chicago in Illinois. The burning of sage "cleanses the house so there is room to talk objectively", explains Gwin. Now he is ready to talk about — and in — the language of his ancestors.

Hidatsa, like many languages, is on the verge of vanishing and taking with it crucial linguistic and cultural data. As fluent speakers grow older and major languages such as English, Spanish, Arabic and Mandarin Chinese overwhelm small cultures, fewer young people choose to cultivate their native tongue. Only about 75 people speak Hidatsa fluently. Most of them, including Gwin, are over 50.

The world contains about 6,900 languages, but linguists estimate that at least half of these will vanish during the next century. Languages are fluid systems, constantly changing and adapting to speakers' needs; death is a natural part of that process. Yet languages are disappearing at an unprecedented rate. Every ten days or so, the last fluent speaker of a language dies, erasing key linguistic information. "Losing languages is bad for science," says Boyle.

Fragments of some are retained in written documents or recordings. But most have never been written down. They vanish without any documentation of their sounds, words or sentence structure. Such information can provide vital clues to understanding how the brain acquires, organizes and processes language.

In addition, small and endangered languages often display rare characteristics that help linguists understand the limits and versatility of language. They may harbour knowledge about the natural world and even offer insight into human migrations.

Strong language

Yet linguists have not always appreciated the importance of rare tongues. In the 1950s and 1960s, the influential linguist Noam Chomsky proposed that the human brain is prewired to learn language¹. His research prompted others to search for language 'universals' that underpin all tongues and so offer insight into the building blocks of human thought.

Some field researchers continued to document the quirky characteristics of little-known languages, which often challenged the language universals. "Essentially, every time we find another language, another universal bites the dust," says Doug Whalen, a linguist at Haskins Laboratories in New Haven, Connecticut.

For example, an important language universal involves the order of words in a sentence. An English sentence such as 'the boy hit the ball' follows a subject-verb-object (SVO)

word order. In Hidatsa, the same sentence would read '*maagarishdawacee ma'uudabi nigic*', or 'the boy the ball hit' — a pattern known as SOV. These are the two most common patterns, and certain others were long thought to be impossible.

But in the 1980s, linguists studying rare tongues in the Amazon, discovered the object-verb-subject (OVS) word order, which translates literally as 'the ball hit the boy'². "If linguists hadn't noticed those languages," says David Harrison of Swarthmore College in Pennsylvania, "we might still have a mistaken idea that OVS is an impossible structure for a language."

By the 1990s, linguists had joined forces to voice their concern about language loss³. This spring, the US National Science Foundation and National Endowment for the Humanities established the \$4.4-million Documenting Endangered Languages project. Another non-profit group, the Endangered Language Fund, is supporting Boyle's Hidatsa work.

On a September afternoon a few miles east of the North Dakota badlands, Boyle hunches over a notebook and elicits phrases from Gwin. "The part I like is when John throws crazy sentences at me," says Gwin, grinning.

Gwin calls his language a tool handed down from his grandmother and mother. "Native languages go to where the English language cannot travel," he says. "They are the key to talking to the spiritual world." Hidatsa was not always treated with such reverence; as a girl, Gwin's grandmother, Pearl Burr Young Bear, was placed in a boarding school where English was the only language permitted. Pearl would sneak into the boiler rooms with her friends to practise their Native American languages.

On this trip, Boyle is working with Gwin to

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

find out how Hidatsa forms subordinate clauses, which in English often involve words such as 'when,' 'if' and 'because'. He is also studying how Hidatsa coordinates nouns, noun phrases and clauses.

For the latter, Boyle asks questions such as "How do you say 'the man and the girl sing'?" At one point he accidentally asks, "How do you say 'the man can dance and sing'?" Gwin spoke the sentence in Hidatsa; Boyle, surprised, asked him to repeat it.

"I didn't even think about how I was asking it," says Boyle, but that small slip of the tongue revealed a characteristic of Hidatsa that he didn't know: the use of 'can' as a modal verb. Linguists had documented this in Crow, the language most closely related to Hidatsa, but never in Hidatsa itself. "Eliciting artificial sentences tells you a lot about the grammar and structure of the language," Boyle says.

The written word

Field linguists often work for years with the goal of producing a dictionary of 10,000 to 15,000 words, a 300- to 400-page grammar describing the language, and a group of texts that show how the language is used. So far, Boyle has a basic Hidatsa dictionary of 4,500 nouns and verbs, a basic grammar book designed for high-school students, and a book of 133 irregular verbs. Native speakers use materials such as these in revitalization programmes; linguists study the details to define the limits of human cognition and language diversity.

For instance, many languages use suffixes called illocutionary markers to define the truth value of a statement. Hidatsa and other languages in the Siouan language family stand out by having as many as 18 of these markers.

In Hidatsa, for example, the basic sentence 'the man kissed the woman' is built using the words *macée* for man, *wiá* for woman and *iigiracóobi* for kiss. But a Hidatsa speaker would have to explain how he or she came upon this information. If the speaker witnessed the event and knows it for a fact, then markers would be added to make the sentence read '*macéeš wiáha iigiracóobitooreš*'. But if the speaker is telling a traditional story passed down through generations, the final marker would change, making the sentence: '*macéeš wiáha iigiracóobiwarec*'.

Documenting such variability shows the many ways in which people can put words together, explains Alice Harris, a linguist at the State University of New York in Stony Brook. "Languages are the natural laboratory for linguists," she says.

At Swarthmore College, Harrison has spent years documenting rare languages in Siberia. He discovered that one such language, Tofa, has a single suffix, *-sig*, that can be added to any noun, changing it into a word that means 'smelling of' or 'smelling like'. For instance, the word for reindeer in Tofa is *ivi*, so *ivisig* means 'smelling like a reindeer'. The smell suffix had never been reported before; why it is impor-

tant to the Tofa people remains a mystery. But linguists argue that little pieces of information like this are essential for understanding the limits of how the brain organizes language.

"If linguists had only major world languages to study — say, Japanese, Hindi and Spanish — we would be severely handicapped in understanding human cognition," explains Harrison. "Linguists need the oddest, quirkiest and most unusual languages and words to test our theoretical models."

Languages of small cultures can even shed light on human migration. "A language is probably the most important artefact a culture possesses," says Robert Rankin, a professor emeritus at the University of Kansas. "Language can tell us about the things that archaeology cannot."

Speaking volumes

For example, the material artefacts of the Wiyot and Yurok tribes of northern California bear little resemblance to those of the Ojibwa, Shawnee and similar tribes of the northeast. But their languages share essential characteristics of grammar and vocabulary. "So the tribes must have come from the same group at some time in the distant past," says Rankin. "This is really important information that archaeology could never have revealed."

Small and endangered languages can also harbour indigenous knowledge, for example about medicinal plants and pesticides. This knowledge represents long-term adaptations to the land, says Harrison, and often cannot be easily transferred into another language. "Small cultures are a repository of knowledge about nature that is about to be lost," he says.

In Tofa, for instance, each month is named after a hunting or gathering activity. The word for May means 'digging *saranki* root month' because it is when locals collect the bulb of the lily-like *saranki* flower, to be used year-round to treat colds and other illnesses. November is 'hunting month', and July 'hay-cutting month'. The knowledge embedded in these words is lost when people begin using a more common language. Tofa children who now speak Russian no longer retain the monthly information, and many elders have also forgotten it.

Activists hope that the new push to save endangered languages will make a difference, but it is unclear whether the efforts will stem the rate of language loss. For many, time is running out. Nearly 550 languages have fewer than 100 fluent speakers.

And even as Gwin teaches Hidatsa in the schools, most children on the reservation speak English as their first language. Hidatsa may survive another generation or two, but ultimately it, too, is likely to vanish into the pages of history.

Jessica Ebert is a freelance writer in Minnesota.



John Boyle (top right) talks Hidatsa with Alex Gwin. Sergei Kongarayev (bottom) is only partly fluent in the Siberian tongue Tofa.

1. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, Cambridge, Massachusetts, 1965).
2. Olawsky, K. J. *Leiden Pap. Linguist.* **2**, 43–68 (2005).
3. Hale, K. et al. *Language* **68**, 1–42 (1992).

All in the mind of a mouse

Could mice with faulty genes help us to understand the biology of psychiatric disease? **Carina Dennis** investigates.

First, there is the cowering in the darkness. Then the furtive scurrying and crouching against the wall. But what really grabs the attention of the former CIA agent is the nervous scratching. An expert in video surveillance, he is programming a computer to monitor these movements. But, unlike in his previous work, he is not tracking suspicious behaviour that could flag a possible terrorist or criminal. This time his attention is focused on the antics of a mouse.

The expert, who won't reveal his identity, is collaborating with geneticist Mario Capecchi at the University of Utah in Salt Lake City to develop computer technology that will aid studies of psychiatric disorders in mice. The aim is to let researchers screen the behaviour of large numbers of mice without having to spend thousands of hours glued to a video monitor. Scientists hope that such developments will let them unleash the full power of mouse genetics on the challenging problem of human psychiatric disease.

There is little doubt that genes play a significant role in psychiatric disorders; a raft of genes has been implicated in conditions such as depression and schizophrenia^{1,2}. But tracking down these genes can be extremely difficult. And there is much to be learned about how the culprit genes that have been identified contribute to the onset and progression of the disorders.

Studying the faulty genes is a challenge. For one thing, it is nearly impossible to do the kinds of genetic and molecular experiment needed to work out how these mutant genes cause disease in people. For another, researchers need good animal models in order to test and develop new drugs. "The big change for the field is being able to model the genetic aspects of psychiatric illness in mice," says Daniel Weinberger, a

psychiatrist at the National Institute of Mental Health (NIMH) in Bethesda, Maryland, who is using mice to test different forms of a human gene implicated in schizophrenia.

Timid tendencies

The great advantage of mice is that, unlike rats, they can readily be genetically engineered — in this case, to carry a specific gene mutation known, or thought, to cause human disease. The engineered mice can then be rigorously tested to see how the mutant gene affects the animal's behaviour, cognition and physiology, and used to test new therapies. Breeding the mice with other strains can show whether a different genetic make-up influences the effects of the mutation. Their completed genome sequence and rapid reproduction also means mice are easy to screen in large numbers for new disease genes, whose equivalents can then be tracked down in people.

But mice have a big drawback: they are extremely nervous creatures that become stressed when people are near. This makes it hard for scientists to tell whether a behaviour they observe is down to a mutant gene or, say, sheer fright. For that reason, most psychiatric work has, until now, focused on the more non-chalant rat. "The mouse has really been underestimated," says Tim Bussey, a behavioural neuroscientist at the University of Cambridge, UK. "People think it is more difficult to do behaviour studies with mice than with rats, but

it's often easier." Now, new screening technology and more sophisticated behavioural tests are giving the field a boost.

Researchers are using a range of approaches to study how genes influence specific behaviours observed in psychiatric conditions. Some are targeting known genes, whereas others are randomly disrupting genes in the mouse genome and screening the resultant mutants for behavioural changes. Yet others are looking for natural behavioural variations in mouse strains and tracking down the culprit genes.

But once armed with their chosen mice, geneticists face two key problems: the timidity of their subjects, and the question of how to interpret the behaviour they see.

Secret surveillance

To tackle the first problem, researchers have devised ingenious ways to achieve something like a Big Brother mouse house, where the mice can get up to their usual antics unaware that humans are monitoring their every rustle and tussle.

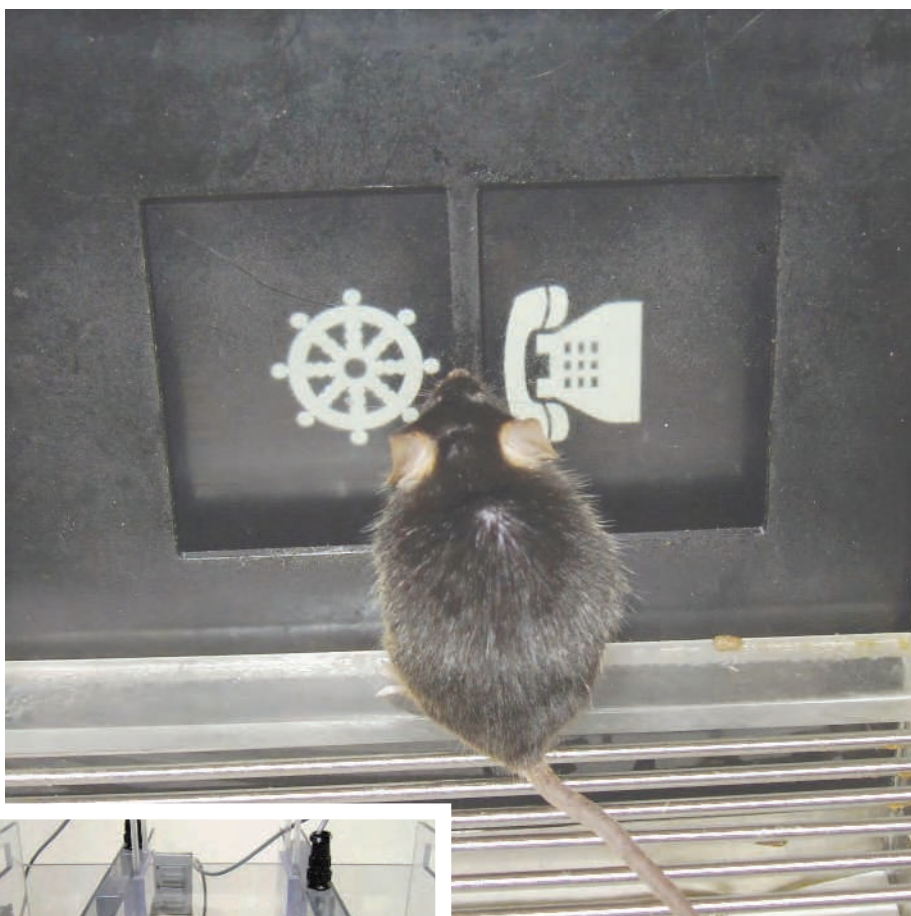
To help Capecchi spy on mice, the CIA computer expert is automating the recording of highly repetitive actions in a mouse model of obsessive compulsive disorder (OCD); that is, the actions of mice that lack their normal *Hoxb8* gene³. "Watching hours and hours of video footage was cumbersome — we needed to automate the process," says Capecchi.

He plans to compare the behaviour of mice with different genetic alterations of *Hoxb8*, which are the same as those found in humans, to see how the different mutations affect mice's actions. He hopes this will help him to identify mutations that cause OCD and related conditions in humans. Capecchi also plans to use the technology to study abnormal social interactions in mice with an altered

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

"Mice are a very social species so they are a good choice for modelling symptoms of autism."

— Jacqueline Crawley



Pause for thought: tests of learning ability (above) and sociability (left) in mutant mice can provide clues about behavioural disorders in humans.

Hoxa1 gene, which is implicated in autism.

Other groups — such as that led by Hans-Peter Lipp at the University of Zürich, Switzerland — have created a 'mouse hotel', which sleeps up to 16 mice and has a wide array of platforms and gadgets to test behaviour and social function. "It doesn't make sense to test mice that are isolated because usually they live in a social environment," says Lipp. A tiny microchip implanted in the back of each mouse's neck allows their movements and interactions to be automatically recorded⁴.

Seriously social

But unobtrusive researchers are still left with the second problem: diagnosing the mouse's behaviour. Although mice are good models for many aspects of human biology, it is unlikely that they could fully mimic all the complexities of a human psychiatric disease such as schizophrenia. Instead, researchers are breaking down these diseases into their simpler components, such as the anxiety involved in depression, or the lack of sociability associated with

autism, and using mice to gain insights into the neurobiology underlying such traits.

"Mice are a very social species so they are a good choice for modelling symptoms of autism," says Jacqueline Crawley, who heads the Laboratory of Behavioral Neuroscience at the NIMH. She developed her mouse sociability tests after watching autistic children interact at a clinic. Her tests involve a three-chambered apparatus (pictured), and each mouse is able to choose to spend time with a novel object or with a stranger mouse^{5,6}. Normal mice would rather get acquainted with the stranger, but Crawley and her team have identified several inbred strains that, like autistic children, fail to show sociability.

Geneticist Richard Paylor, of the Baylor College of Medicine in Houston, Texas, and his team are also homing in on autism. Their approach is to study mice engineered to have a mutation in the gene *Fmr1*. Humans who don't express this gene suffer from a mental retardation condition called fragile-X syndrome, and up to 20% of patients with the syndrome also have autism. Paylor wants to understand why, when all the patients lack the same gene function, only some develop autism. He and his team have devised an experiment that tests whether a mouse's willingness to socialize stops when it is presented with a new mouse or

a new environment; difficulties in dealing with new social situations is a characteristic of fragile X in humans.

By crossing the mutant mice with different strains, Paylor and his colleagues have shown that some of the offspring avoid all social interaction, more typical of autism. They hope to identify the genes that exacerbate the fragile-X traits towards more severe autistic behaviour.

From mouse to man

Teasing out these crucial shifts in behaviour often requires subtle tests. Seth Grant, a geneticist from the Sanger Centre in Cambridge, UK, for example, is investigating how a gene encoding a protein called SAP102 contributes to learning and behaviour. This gene is mutated in humans with mental retardation⁷, but mice lacking SAP102 have less obvious defects. With training, the mice can overcome their learning deficit, but will keep on using the same strategy even when an easier option is available. So, although the animals can learn, they are relatively inflexible in the choices they make.

But how far can you compare, say, anxiety in a mouse to anxiety in a human? Some researchers go as far as applying the same behavioural tests to humans and mice. Bussey, for example, has taken a standard touchscreen test used to assess cognitive function in humans and adapted it to mice. The mouse is shown two images on a screen (pictured) and learns that if it touches one, it gets a food pellet. Then experimenters reverse the rules to see how quickly the animal can adapt to change, an ability that is impaired in some human behavioural disorders.

Mark Geyer, a psychopharmacologist at the University of California, San Diego, has turned the tables and applied a mouse test for exploratory behaviour — called an open-field test — to human patients. He has designed an environment for patients, in this case a bogus 'office', where they are asked to wait. Manic patients behave a lot like manic mice — hyperactively exploring the new environment. According to Geyer, these unpublished data will help to validate his mouse model, which, he hopes, will shed light on the biology of mania and schizophrenia.

So although humans are clearly more complex than mice, perhaps the basic components of our behaviour are more similar than we would like to think. We might not have too long to wait before researchers can use mice to get inside our heads.

Carina Dennis is Nature's Australasian correspondent.

1. Lesch, K. P. *Rev. Psychiatry Neurosci.* **29**, 174–184 (2004).
2. Harrison, P. J. & Weinberger, D. R. *Mol. Psychiatry* **10**, 40–68 (2005).
3. Greer, J. M. & Capocchi, M. R. *Neuron* **33**, 23–34 (2002).
4. Galsworthy, M. J. et al. *Behav. Brain Res.* **157**, 211–217 (2005).
5. Moy, S. S. et al. *Genes Brain Behav.* **3**, 287–302 (2004).
6. Nadler, J. J. et al. *Genes Brain Behav.* **3**, 303–314 (2004).
7. Tarpey, P. et al. *Am. J. Hum. Genet.* **75**, 318–324 (2004).

L SKILLINGS

S. TOLL & J. CRAWLEY, NIMH

BUSINESS

Path to approval proves rocky for copycat biodrugs

Attempts to copy the first generation of biotechnology drugs are facing fierce resistance, as **Meredith Wadman** reports.

Producers of generic drugs are embarking on an epic battle to win regulatory approval for the first copies of the complex biological drugs that underpin the biotechnology industry.

Next week, a European Union law comes into effect that will define a pathway to the market for generic versions of the drugs, and the European Medicines Agency is expected to give its first approvals of such drugs sometime next year.

These could include a copy of human growth hormone and the hepatitis C treatment interferon- α , made by Swiss company BioPartners. Another possible candidate is a version of California-based Amgen's blockbuster anaemia treatment erythropoietin, made by Croatian drug company Pliva.

But the prospect of biogenerics is viewed by the biotechnology industry as a significant threat now that the patents on the founding generation of its products are expiring. And the industry is using every argument it can find — including possible safety concerns — to press regulators for high hurdles to their approval.

"We believe in biosimilars," says Roger Perlmutter, Amgen's head of research, using a term the industry prefers for generic biopharmaceuticals. "But from a public-health point of view, you don't want to put molecules on the market that are less safe" than the ones they're meant to imitate.

The makers of generics counter-charge that the industry is using scare tactics to protect lucrative markets. "There is no reason to say that these products are any more risky" than standard drugs, says Andreas Rummelt, chief executive of Sandoz, the generics division of Novartis and one of the world's top two generics manufacturers. "Patients and physicians

need to have access to safe and effective follow-on products once patents expire."

But as a regulatory path forward opens in Europe, its US equivalent remains elusive. The Food and Drug Administration (FDA) seems to have stalled in its efforts to find one — prompting Sandoz to sue it a couple of months ago. The company is demanding action on its 28-month-old application to bring a generic version of human growth hormone to market (see *Nature Rev. Drug Discov.* **4**, 798–799; 2005).

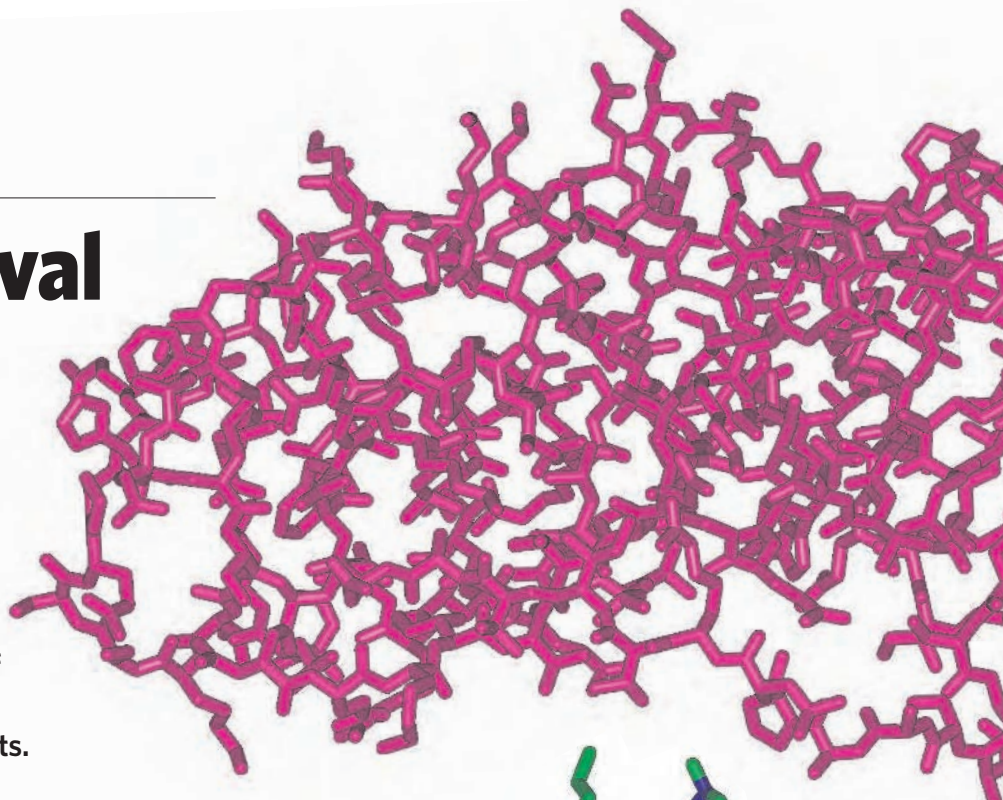
"Scientists at the agency have said the application can be approved," asserts Eric Pomerantz, general counsel at Sandoz. "There are no technical, scientific, regulatory or legal reasons for not approving it."

Plan goes slow

Almost three years ago, then-FDA commissioner Mark McClellan pledged to move ahead with biogenerics, but this hasn't happened yet. Helen Winkle, director of the agency's Office of Pharmaceutical Science, told a generics industry meeting in Washington DC last month that the agency is "struggling" with the issues raised by biogenerics approvals.

It is not hard to see why. The distinctive feature of biological drugs is their size and complexity. Most are large proteins with appended sugar groups — such as erythropoietin and the multiple sclerosis treatment interferon- β . They are typically 100 to 1,000 times larger than traditional drugs (see above), and are produced in cultured mammalian cell lines, with their inherent variability, rather than in test tubes.

Because of the compounds' chemical and biological complexity, their purification can involve hundreds of steps, rather than the couple of dozen often required for conventional drugs. Furthermore, almost all



Little and large: biological drugs such as erythropoietin (purple), used to treat anaemia, are far larger and more complex than standard drugs such as the impotence treatment sildenafil (Viagra).

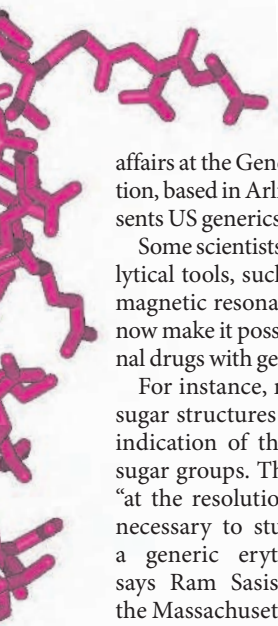
biological compounds may provoke rare but serious immune reactions. This immunogenicity adds another layer to the safety assessment required for both innovative and imitative biological drugs.

That has strengthened the hand of the biotechnology industry, which is arguing in the United States that generics makers should be required to go through many of the costly clinical trials demanded of innovator companies when they first bring biological drugs onto the market.

In the case of traditional chemical compounds, "the generic drug is the same as the original, and will behave the same way as the original", argues Amgen's Perlmutter. "You can't do that with biosimilars. Every product is different. And you must test that product [in people] to ensure efficacy and safety. You can't assess that product by any physical test."

The generics industry hotly disputes this last contention, and is calling for a "flexible, abbreviated, approval process" similar to the speedier approval process created for standard generic drugs in the United States by a 1984 law. "An abbreviated approval process is clearly within the scope of current science," says Gordon Johnston, vice-president for regulatory

R. SASSEKHARAN



affairs at the Generic Pharmaceutical Association, based in Arlington, Virginia, which represents US generics makers.

Some scientists concur, saying that new analytical tools, such as high-resolution nuclear magnetic resonance and mass spectrometry, now make it possible to closely compare original drugs with generics, at least in some cases.

For instance, mass spectrometry can map sugar structures on proteins, giving a better indication of the locations and identity of sugar groups. This mapping is now possible "at the resolution that the FDA would feel necessary to study the difference between a generic erythropoietin and Amgen's", says Ram Sasisekharan, a bioengineer at the Massachusetts Institute of Technology in Cambridge, who has studied the feasibility of such comparisons.

And just as biological drugs vary greatly, so too does the range of analytical tools available to establish the bioequivalence of a generic to its original patented analogue. Thus it is widely accepted that the regulatory requirements for biogenerics will have to be assessed on a case-by-case basis.

Money matters

But if these requirements become too onerous, generics makers may walk away. Already the cost, risk and skill required to enter the area makes biogenerics a challenging target for all but a few companies, says David Evans of Data-monitor, a London-based business information company.

There is one other thing that all sides agree on: plenty of money is at stake. For instance, Pfizer, the original maker of human growth hormone, had \$239 million in US sales of the drug last year according to data firm IMS Health — even though the key patent on the drug has long since expired. Those sales could be seriously eroded if the US District Court rules in favour of Sandoz in the current lawsuit, and the FDA approves Sandoz's generic version.

"This is a huge new opportunity for generics companies — especially as generics growth is slowing in mature markets such as the United States," says Evans. He estimates that the market for drugs that could be replaced by biogenerics was worth \$20 billion last year.

In the meantime, US regulators will have to determine how to handle approval applications. "A lot of these molecules are going off patent," says Sasisekharan. "This is something we need to deal with. We can't avoid it. We can't postpone it. The most important issue is to figure out the path forward."

IN BRIEF

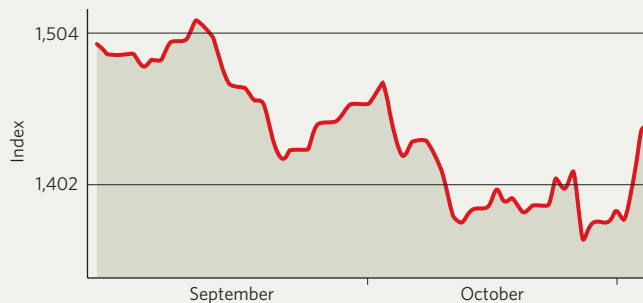
VIOXX VERDICT 2 Merck scored a major victory in the second US lawsuit brought over the safety of its painkiller Vioxx. A nine-member jury in a New Jersey court found on 3 November that Merck had fairly marketed and fairly represented the risks of the painkiller. Vioxx was withdrawn in September 2004 after it was found that prolonged use increased the risk of heart attack and stroke. The jury concluded that the drug did not cause the heart attack of Frederick Humeston, a 60-year-old postal worker, in 2001. Merck lost the first Vioxx lawsuit in Texas in August (see *Nature* 436, 1070; 2005). Its stock increased by 4% to \$29.48 on the latest news.

HIRING BINGE Samsung says it will employ an extra 26,000 researchers in the next five years as part of its drive to consolidate its position as one of the world's largest electronics firms. Lee Yoon-woo, chief technology officer of the Korean company, said that he would double the research and development workforce to 52,000 as the company moved to secure its expansion into areas such as mobile phones and laptop computers. He made his comments at a jamboree on 3–4 November, held for several hundred financial analysts, to discuss Samsung's expansion plans.

CASH DASH The average total pay package for research and development chiefs of US biotechnology companies jumped by 12% in 2004, according to an analysis conducted by *BioWorld Today*, a biotechnology newspaper based in Atlanta, Georgia. Vice-presidents of research and development at 145 companies averaged \$366,000 in salary plus bonuses, up from \$326,000 in 2003. Total pay for chief executives rose 5%, to \$596,000, whereas that for chief legal officers at biotech firms dropped a little, to \$242,000.

MARKET WATCH

Nanotechnology stocks



Nanotechnology stocks have taken a tumble this autumn, with a couple of particularly troubled stocks compounding the problems of a generally weak market in technology stocks.

The Lux Nanotech Index, which comprises companies supplying nanotechnology products as well as some major industrial companies that rely heavily on nanotechnology, dipped close to its lowest point since 2003 at the end of last month.

"The index has been down, like the broader market," says Peter Hebert of New York-based Lux Research, which compiles the index, adding that he thinks it will rise again soon. Hebert's company has just launched a fund called the PowerShares Lux Nanotech Portfolio, which will allow investors to track the performance of the index.

The weakest performers over the past two months included Skyepharma, a supplier of nanotech-based drug delivery systems based in London, and Headwaters, whose main subsidiary, Nanokinetix of New Jersey, makes catalyst components.

Skyepharma announced "disastrous" first-half results in September, Hebert says, and a share offering failed to attract buyers at the anticipated price, causing its shares to dive. Headwaters fell back sharply from a short-lived summer peak attained on a good set of financial results.

Hebert predicts that stock in the companies represented by the index will move up again "as nanotechnology becomes a significant component of their businesses", and says that subdued stock prices make this a good time to get into the new fund.

Colin Macilwain

LUX RES.

Biodiversity needs the help of global change managers, not museum-keepers

SIR — As observed in your News Feature “Dollars and sense” (*Nature* 437, 614–616; 2005), there is increasing evidence that many conservation organizations remain focused on single species instead of addressing the urgent problems caused by loss of ecosystem functionality.

More and more conservation scientists are calling for a holistic approach that considers ecological processes and the functional properties of ecosystems, rather than just parts and patterns of species, as crucial conservation targets (see, for example, P. Kareiva and M. Marvier *Am. Sci.* 91, 344–351; 2003).

However, ecosystem functions — which become services when used by people — are not yet considered in most mainstream conservation approaches. Another shortcoming is the rather static view of biodiversity held by many conservation organizations today. Thus, there are even more reasons for a paradigm shift in conservation than those addressed in your News Feature.

“The anthropogenic climate change that is expected during the next century looms as an overarching and unprecedented threat.”

— P. Ibisch, M. Jennings, S. Kreft

Ecosystem functionality means that an ecosystem itself can sustain processes required to maintain its parts by being, for example, resilient enough to return to its previous state after environmental disturbance. Functionality depends in various specific ways on the quantity and quality of a system's biodiversity. An important characteristic of ecosystem functionality is that it develops and responds dynamically to constantly occurring environmental changes.

The anthropogenic climate change that is expected during the next century looms as an overarching and unprecedented threat to biodiversity. The predicted rate of warming alone may move many species well beyond their current climate-niche ranges.

Some species will find themselves in habitats that are unsuitable in many secondary ways, for example, as specific breeding microhabitats or for symbiotic interaction with other species.

Further, individual species within an ecosystem will be threatened by unpredictable factors, such as changes in seasonal resources and in the biogeography

of pathogens, predators and competitors, which could trigger extinction events.

Although ecosystems never have been in a steady state and species distributions have always been on the move at one timescale or another, it is now more clear than ever that it is impossible to statically conserve current biodiversity patterns, in hotspots or anywhere else.

Unfortunately, many conservationists have not yet grasped the need to be ‘global change managers’ rather than museum-keepers, a shift of perception that is urgently required to mitigate the impacts of global change and help ecosystems adapt to them.

Pierre L. Ibisch*, **Michael D. Jennings†**, **Stefan Kreft***

*University of Applied Sciences Eberswalde, Alfred-Möller-Str. 1, 16225 Eberswalde, Germany

†The Nature Conservancy, Global Conservation Approach Team, 530 South Asbury, Moscow, Idaho 83843, USA

Biodiversity: journals must take a broader view

SIR — Biodiversity hotspots have been useful tools in prioritization, particularly in identification of critical gaps in protected areas. The analysis of avian hotspots cited in your News Feature “Dollars and sense” (*Nature* 437, 614–616; 2005) did not find them ineffective, but recognized that hotspots based on different taxa or indices are not necessarily congruent and a synthetic approach is required.

Although, as your News Feature suggests, more attention needs to be paid to preserving ecosystem function, we are facing a biodiversity crisis. Neither ecosystem function nor hotspots should be the sole focus of conservation efforts: we need both. Arguing the economic perspective may be a good approach to lobbying, but it is not a replacement for urgent, targeted action.

Conservation efforts also require evaluation: audits require detailed appraisal, in addition to any reporting required by donors. Large conservation organizations can fulfil these criteria relatively easily, but most conservation practitioners are small scale, depending on volunteers, drawing on very limited funds and lacking spare capacity to permit such audit. Limited audit would provide a certain amount of information, but only from the most easily reviewed and most positive cases.

For conservation efforts to be maximally useful, failures must be reported as candidly as successes.

The real gap lies not so much in analysis but in reporting: we need journal editors to take a broader view of what is of interest to a wide readership and to consider more

case studies, even when these are not ‘groundbreaking’. Publication of results needs to be brought into the mainstream. This requires a major editorial change, allowing a shift away from the current domination of analysis and theory, to reporting of real conservation science.

Justin Gerlach

Nature Protection Trust of Seychelles, University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, UK

Biodiversity: saving Florida panther makes sense

SIR — Your News Feature “Dollars and Sense” (*Nature* 437, 614–616; 2005) asks whether we should focus more on economic value and less on the biological needs of imperilled species. You give the Florida panther (*Puma concolor coryi*) as an example of endangered species recovery that may not make “economic or scientific sense”. In fact, it does make sense.

The Florida panther was listed by the Endangered Species Act (ESA) in 1973, when perhaps as few as 30 individuals remained in south Florida. The population contained low genetic variation and physical abnormalities associated with inbreeding depression.

In 1995, wildlife managers embarked on a genetic restoration programme, releasing female Texas pumas into south Florida. A subsequent reduction in genetically based defects and an increase in survival and reproduction suggest that the programme was a success.

Today the population numbers nearly 90 individuals, an astonishing increase, directly attributable to ESA measures. Although significant threats remain, the panther now has a fighting chance at recovery.

The News Feature does not consider the economic value of conserving panthers. The species is a major attraction for tourists, and more than 1.3 million speciality licence plates have been purchased by people in Florida, generating more than \$30 million for panther conservation.

Interested readers may contact the author at hartt@nwf.org for a list of publications on Florida panther conservation.

Laura Hartt

National Wildlife Federation, 1330 West Peachtree Street, Suite 475, Atlanta, Georgia 30309, USA

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

AUTUMN BOOKS

Peaks in climate research

Lonnie Thompson climbs every mountain to look for clues to climate change.

Thin Ice: Unlocking the Secrets of Climate in the World's Highest Mountains

by Mark Bowen

Henry Holt: 2005. 480 pp. \$30

Georg Hoffmann

"The closest living thing to Indiana Jones, and just in time." That's how Harvard geochemist Daniel Schrag described Lonnie Thompson, the pioneer of high-altitude ice-core drilling of tropical and subtropical glaciers. Thompson has probably spent more time at altitudes of 6,000 metres and above than anyone else. But when he started his drilling campaigns in the 1970s, in such remote places as the Quelccaya ice sheet in Peru, it was an adventure against long odds. Starting from his laboratory at Ohio State University, Thompson faced technical, practical and scientific problems considered by most of his colleagues to be unsolvable. More than thirty years later, the first summary of this most unusual career has appeared.

In *Thin Ice*, Mark Bowen, a physicist and passionate mountain-climber, describes the numerous expeditions that Thompson and his colleagues organized — from the Andean Altiplano to Tibetan ice sheets, from the Alaskan Bona-Churchill glacier to Mount Kilimanjaro in Kenya. Of the peaks capped with drillable amounts of ice, only a handful still await exploration by Thompson and his group.

It was widely believed that even if the researchers overcame the impossibility of retrieving an ice core from high altitudes, they would never be able to date, analyse and interpret the ice appropriately. In fact, they demonstrated that the entire concept was entirely feasible. When attempts to airlift heavy polar drills to high-altitude glaciers failed, they designed and constructed special lightweight drilling equipment that could be carried on foot. But these technical advances did not spare Thompson's team the discomforts of draughty, cold, low-oxygen conditions atop the glaciers, on which they spent months locating high-quality ice samples and battling with a range of technical problems.

Chemical and isotopic analyses of tropical ice cores showed convincingly that the tropics, far from being an immutable part of the global climate system, were participating in the rapid cold and warm swings documented at higher latitudes. The publication of the results from Huascarán and Sajama, two Andean mountain



giants that Thompson successfully drilled, changed many climatologists' way of thinking. Bowen has adopted the still controversial idea that the tropics are a significant driver of rapid climate variability, arguing against what he calls the North Atlantic school. But most palaeoclimatologists still belong to the latter camp, remaining convinced that the most potent regulator of the global climate system is hidden somewhere near Greenland at the turnover point of the ocean's thermohaline circulation. Even though I am personally not yet convinced by the arguments of the 'tropical school', I found this vivid description of an ongoing scientific discussion to be one of the most compelling stories that *Thin Ice* has to tell.

Bowen adds a political dimension to the scientific narrative, arguing strongly for the promotion of climate change to the ranks of the world's most pressing problems. The assertion is justified by the race against time for researchers such as Thompson who are trying to extract information about past and present climate shifts from the world's disappearing glaciers. Had Thompson started his work today, many of the frozen archives he used would not have been considered as potentially

rewarding drilling sites — the quality of many signals in the ice has already been degraded by recent warming. The Quelccaya ice sheet, for example, no longer experiences surface temperatures low enough to allow the preservation of a seasonal water isotope signal. Meltwater is now percolating down through the upper metres of compacted snow, affecting all proxy information buried in the ice. Quelccaya's glaciers have become hundreds of metres shorter since Thompson started his work thirty years ago.

We can only speculate on how the disappearing tropical glaciers and their water resources will affect those living in Peru and Bolivia in the near future,

but we have some good guesses about what regional climate swings meant to the people living on the Altiplano in the past. Thompson's work has been enthusiastically embraced by archaeologists, particularly those interested in the rise and fall of various pre-Incan societies. Sequences of favourable wet periods and adverse dry periods were accurately archived in Quelccaya's ice, and provide a convincing explanation for the evolution of ancient civilizations of the New World.

Bowen, who accompanied Thompson on two of his expeditions, tells this story in an amusing and entertaining style. The book often changes its focus, switching from thrilling mountaineering anecdotes to interviews with different scientists and historical retrospectives of Earth science. Often Bowen delivers fine pieces of science journalism, describing for example Charles Keeling's first measurements of carbon dioxide in the 1950s on Mauna Loa, Hawaii. *Thin Ice* is at the same time a scientific biography, a fine introduction to the sciences of climate change, and a vivid description of a geophysicist's work under most extreme conditions. ■

Georg Hoffmann is at LSCE/CEA, Orme des Merisiers, Gif sur Yvette 91191, France.

ILLUSTRATIONS BY CHRISTIAN DARKIN

Magnetic personalities

Fatal Attraction: Magnetic Mysteries and the Enlightenment
by Patricia Fara

Icon Books: 2005. 204 pp. £9.99

David W. Hughes

It is a great shame that the history of physics is not an integral part of a standard university physics education. Teaching it would provide an ideal means of dragging the laboratory- and computer-based academic physicist out into the wider world. Not only would it underline the contemporary relevance of physics, it would also civilize the subject by reconnecting it with its roots.

One requirement would be a collection of excellent introductory books on the history of physics — and Patricia Fara's *Fatal Attraction* is a trail-blazer for magnetism. Fara, a lecturer in the history and philosophy of science at the University of Cambridge, accurately targets the enquiring student reader. The writing is pacy, informative, riveting and unencumbered by footnotes. But the book also contains a copious reference section and draws extensively from Fara's scholarly tome *Sympathetic Attractions* (Princeton University Press, 1996).

Magnetism has a past rooted in the mists of antiquity. Certain types of magnetite iron oxide ore, when struck by lightning, become strongly magnetized. This natural loadstone was once greatly treasured and was first used by the Chinese in the fourth century BC to produce the magnetic compass, which found application both in the art of feng-shui and as a navigational device.

Fara takes us back to the times before magnetism was allied with electricity. Her main period is the Enlightenment in the seventeenth and eighteenth century, when natural philosophers and thinkers optimistically believed that the problems of society could be solved by reason and common sense. In those days, cosmic magnetic forces were the precursors of universal gravity. England's wealth and protection depended on sailors successfully navigating rough oceans, so the magnetic compass was of paramount importance. And the dawn of today's alternative medicine was heralded by the supposed hidden powers of magnets. People were happy to strap magnets to the afflicted parts of their bodies, expecting them to draw out the pain and heal toothache, rheumatism, gout, scurvy and a host of other ailments.

The story is built around three main characters: Edmond Halley, Gowin Knight and Franz Anton Mesmer.

Halley, who in 1720 became England's second Astronomer Royal, was not only a scientific polymath but also a naval adventurer who sailed the Atlantic measuring magnetic variation (the difference between true and magnetic



north) in the hope of finding a way to calculate longitude at sea. He was the first scientist to insist that the government pay for his instrumentation and finance a scientific mission. Halley was a great believer in speedy publication and in sharing his results (especially as they were obtained by a paid public servant).

The charts of magnetic variation drawn up by Halley were revolutionary, as he used lines to link places of equal variation. These 'halleyan lines' (now known as isogonics) have become a ubiquitous feature of modern meteorological charts. Spurred on by William Gilbert's textbook on geomagnetism from 1600

and Henry Gellibrand's discovery in 1635 that the magnetic variation changed with time, Halley proposed that the Earth had four magnetic poles: two on the surface and two on an inner sphere, 500 miles below, that rotated at a slightly different speed.

Gowin Knight was a scientific entrepreneur from the same period whose main aim was to become rich and famous. He pioneered the production of strong artificial steel magnets and an efficient gimballed maritime air compass. He also convinced the state that science mattered — a legacy that still benefits us all.

Franz Mesmer, a Viennese physician who treated Mozart, suggested that the source of the magnetic force was a stream of invisible, weightless cartesian effluvia coursing through the minute pores of the affected substance. This cosmic magnetic fluid was not only the key to our physical well-being, he claimed, but also the

modus operandi of astrological influence. Moving to Paris in 1778, Mesmer then creamed off the wealthiest patients, causing such annoyance to competing doctors that he was eventually forced to flee the country.

Fara ends her excellent historical review by briefly relating how the magic of magnetism was somewhat spoiled by Michael Faraday. Britain's first professional scientist founded electromagnetism, one of the cornerstones of the modern consumer society, and now nearly every aspect of it is well understood. ■

David W. Hughes is professor of astronomy, University of Sheffield, Sheffield S3 7RH, UK.

A poisoned reputation

Between Genius and Genocide: The Tragedy of Fritz Haber, Father of Chemical Warfare

by Daniel Charles

Jonathan Cape: 2005. 313 pp. £20

Published as *Master Mind* by Ecco Press in the US (\$24.95).

John Cornwell

Fritz Haber will be forever linked with the Haber-Bosch recipe, which produces plentiful and cheap supplies of ammonia. Yet he was one of the greatest chemists of the twentieth century and was the founding father of the industrial-military complex. The story of his life, which has all the ingredients of a Thomas Mann novel, is a tale of scientific genius in the service of the German fatherland. Haber was a

Jew who adopted Christianity to advance his career prospects but instead became a victim of the Nazis. He was an outstanding hero of science but stood accused of war crimes.

The process for which Haber is best known is the synthesis of ammonia from two plentiful gases in nature: hydrogen and nitrogen. He collaborated with Carl Bosch to discover, after some 4,000 trials, an ideal catalyst composed of iron and oxides of aluminium, calcium and potassium. The process is virtually unchanged to this day. The discovery culminated in the production of prodigious quantities of artificial nitrogen fertilizer, which allowed the global population to rise to 6 billion from below its estimated upper limit of 3.6 billion. As a result, Haber richly merits consideration as one of the greatest figures of the past millennium.

His fame and reputation, however, have been dogged by controversy and paradox. The same Haber–Bosch process could be used to make high explosives, enabling Germany to prolong the trench warfare of the First World War despite a Royal Naval blockade of mineral nitrogen.

Haber's reputation has been further besmirched by his involvement in poison-gas warfare. On the afternoon of 22 April 1915, Haber and a troop of gas scientists opened the valves of nearly 6,000 cylinders containing chlorine gas in liquid form. A blanket of thick green-yellow gas swept into the Allied trenches, killing some 5,000 Allied troops (a figure that would be revised downwards by Germany after the war) and injuring 10,000 more. The Germans gained only one mile on the Ypres front as a result. Haber was disappointed: he had hoped to unleash a weapon of mass destruction so powerful that it would bring the war to an abrupt end, and he blamed his generals for restricting its use.

These actions earned Haber the reputation of a war criminal, forcing him into hiding for a period after the war. But there was worse. It was Haber who oversaw the research that created the insecticide Zyklon B, which, a decade after this death, would be used in the Nazi gas chambers on his own relatives.

There has long been a need for a book in English on the life of Fritz Haber, and Daniel Charles is the ideal biographer. His previous book, *Lords of the Harvest*, explored biotechnology and the global food supply, an essential dimension of the Haber story.

In *Between Genius and Genocide*, Charles takes in the huge range of Haber's scientific, technological and patriotic interests, including his quest for cheap gold to beat the postwar reparations burden. A typical Haber scheme, hare-brained from one perspective and yet eminently practicable from another, was his attempt in the 1920s to harvest gold from the sea. The Swedish chemist Svante Arrhenius had found tiny amounts of gold in sea water and calculated that every ton of the ocean contained 6 milligrams of the precious metal. After extensive experiments and many trips on the oceans of the world, Haber discovered that Arrhenius was wildly out: sea water contains only 0.01 milligrams of gold per ton.

Nor does Charles neglect the turmoil of Haber's private life, including his marriages. The tale of the suicide of Haber's first wife, Clara — an event that has prompted various contradictory accounts — is a model of scrupulous biography. Clara, one of the first women to get a PhD in Germany, had married Haber in the hope of sharing in his life in science. But it seems she was increasingly isolated by his preoccupation with work, and during the First World War deplored his use of poison gas. After returning from Ypres, Haber threw a party and Clara found him in an "embarrassing situation" with the woman who was to become his second wife. After he



had fallen asleep, she took his service revolver and killed herself.

But Charles's principal focus is Haber as a faustian embodiment of science and technology in the twentieth century. For Charles, Haber's fatal flaw is his willingness "to serve any master who could further his passion for knowledge and progress. He was not an evil man." Charles draws the chilling conclusion that the moral choices that Haber confronted during his life "were not so different from those that we face today".

Charles might have gone further to reflect that underlying Haber's flaws is the proposition, widely accepted and promoted by those involved in the public understanding of science today, that science is morally neutral. The Janus-faced nature of the Haber–Bosch recipe seems to support the contention. In Germany, moreover, scientists traditionally worked under the auspices of a civil service that was

both apolitical and value free. Yet it was precisely this neutrality that provided an alibi for the German scientific community when Jewish researchers were expelled after the Nazis came to power. Haber was forced out of the Kaiser Wilhelm Institute, which he helped to found in Berlin, despite Max Planck's attempt to argue his case with Hitler. "A Jew is a Jew," Hitler shouted.

The Institute for Physical Chemistry in Berlin now bears Haber's name, but this is still controversial. As the historian Fritz Stern comments in his elegant essay on Haber and Einstein: "The memory lives on — dimly in distorted controversy." Charles's admirable biography will elucidate the controversy and shed fresh light on Haber's memory. ■

John Cornwell is director of the Science and Human Dimension Project, Jesus College, Cambridge, UK, and is the author of *Hitler's Scientists*.

Short cut to space-time

A Briefer History of Time

by Stephen Hawking with Leonard Mlodinow

Bantam Books: 2005. 176 pp. \$25

Jim Al-Khalili

The phenomenal success of Stephen Hawking's *A Brief History of Time* demands that the arrival of this new edition be treated as a major publishing event. So let me begin with a few facts. First, *A Briefer History of Time* is not a new book, but rather an updated and reworked edition of the original. Second, it is certainly briefer, at three-quarters the length of the original. Third, the simple black-and-white diagrams of the first book have been

replaced by stylish colour images, ranging from the amusing (Hawking and co-author Leonard Mlodinow strapped into their time machine) to the misleading (space-times, of expanding universes and wormholes, embedded within space-time).

So what is the motivation for *A Briefer History of Time*? I will leave aside any cynical accusation of opportunistic marketing because I believe the authors have made an honest attempt here to rectify what they perceive as a problem with the original: that millions of readers with no scientific background did not get beyond the first chapter before their brains blew up. To remedy this, that first chapter has been chopped into three bite-sized



and hopefully, one imagines, more digestible ones. On the whole I like this, but it does seem a bit of a cheat if readers get through the same amount of material before giving up, only now boasting of having seen off three chapters instead of one.

The new book is certainly easier going. The old third chapter ("The expanding Universe") of 11 in the original is now the seventh chapter of 12, highlighting the additional weighting given to introductory material. The three middle chapters ("Black holes", "Black holes ain't so black" and "The origin and fate of the Universe"), which together made up a total of 70 pages in the original, are now lumped into one chapter just 18 pages long. Elsewhere, every attempt has been made to clarify those passages deemed to be hard going. Finally, out goes the chapter on the arrows of time, the diagrams of light cones and event horizons, and discussions of chaotic boundary conditions, and in comes a new crowd-pleasing chapter on time machines.

I find myself unconvinced by this valiant effort, however. Clearly, the incredible success of *A Brief History of Time* was due to a combination of timing, marketing and the persona of the author. It can never be repeated. But what is often overlooked is that its major, paradoxical attraction was its charming incomprehensibility to the non-physicist — the idea that anyone could take a peak inside one of the greatest minds in science. This is lost in the new book. For millions of people around the world, *A Brief History of Time* would have been the only science book they have ever read or attempted to read. But with the briefer version, I feel the baby has been thrown out with the bathwater. It is just another run-of-the-mill popular science book on modern physics. The topics that it claims to treat more carefully have been covered better elsewhere. In any case, many of the topics left in and flagged as more introductory are just as baffling, abstract and abstruse to non-scientists as those left out. Just because quantum mechanics and

the special theory of relativity are not at the cutting edge of current thinking doesn't mean they are any less counter-intuitive. The two-slit experiment and the notion of the relativity of simultaneity could have been explained

Science in society

Victory and Vexation in Science: Einstein, Bohr, Heisenberg and Others

by Gerald Holton

Harvard University Press: 2005. 244 pp.
\$35, £22.95

Daniel J. Kevles

In *Victory and Vexation in Science*, Gerald Holton, a physicist and historian of science at Harvard University, provides a series of illuminating historical and biographical essays on science and scientists in the twentieth century. This thought-provoking book mixes reminiscence with scholarly reflection, drawing on Holton's deep knowledge of scientists and their intellectual, religious and social engagements.

The 14 essays range over a variety of topics and are organized into two sections: 'Scientists' and 'Science in context'. The first part covers, in addition to the icons in the book's subtitle, the physicists Enrico Fermi, Percy Bridgman and Isidor Isaac Rabi, and the psychologist B. F. Skinner. The subjects in the second section include innovation in science and art, policy for basic science, postmodernism and science, and women in science. The subjects are disparate, but several arresting topics appear and reappear in the volume.

Among them is the religious impulse that Holton finds behind the science of Einstein and Rabi. As a youth, Einstein was deeply religious in some profound non-sectarian sense, even though he was raised in an irreligious household. After the age of 12, when he began encountering science, his religious inclination

better, especially as the latter receives just a single paragraph.

Fresh material, based on advances made over the past two decades, has been included, such as the concept of 'dark energy' and advances in string theory. I am also pleased to see that the discussion of the anthropic principle is retained. This is a hot topic of discussion at the moment in connection with multiverse theories. However, I found it a little surprising that the idea is still treated rather cautiously here.

I have no doubt that *A Briefer History of Time* will soon be on the shelves of every high-street bookstore around the world. This is surely to be welcomed: any book that can reach a wide audience and get across the excitement of science has to be a good thing. And with Hawking enjoying an iconic status not seen in a scientist since Einstein, his role as an ambassador for science should not be underestimated.

Jim Al-Khalili is professor of the public engagement of science in the Department of Physics, University of Surrey, Guildford, Surrey GU2 7XH, UK. He is the author of *Black Holes, Wormholes and Time Machines*.

was transformed into a strongly felt quest to comprehend the physical world. This drive, Holton says, constituted a flight from "personal, everyday life, with all its dreary disappointments, and escape into the world of objective perception and thought". Indeed Einstein once remarked that the tenacious pursuit of a difficult scientific problem demanded "a state of feeling similar to that of a religious person or a lover".

Einstein ultimately embraced a transcendent spiritualism, free of anthropomorphic and what he considered primitive elements. His views irritated the theologian Paul Tillich and angered clerics such as a Roman Catholic cardinal in Boston, who found intimations of atheism in Einstein's theories of space-time. Queried on the point, Einstein declared that he believed in "Spinoza's God, Who concerns Himself in the lawful harmony of the world, not in a God Who concerns Himself with the fate and the doings of mankind".

Unlike Einstein, Rabi was raised as an orthodox Jew, but while he separated from orthodoxy, Holton notes that deep down he remained "God-struck throughout his life". Like Einstein, Rabi saw science as a means of transcendence beyond the visceral concerns of the human species. He once recalled that physics filled him with awe and put him in touch with a sense of original causes. "Whenever one of my students came to me with a scientific project, I asked only one question, 'Will it bring you nearer to God?'"

The role that intuition plays in science is

also discussed. Holton raises the issue in a captivating essay on the origins of the Fermi group's research with slow neutrons in Rome during the 1930s. The decisive experimental step was taken by Fermi himself, when he interposed paraffin between the fast-neutron source and the target. Fermi turned to the paraffin with neither forethought nor announcement. He was guided, Holton writes, by brilliant intuition, a speculative move "below the level of consciousness". In the course of mathematical invention, Henri Poincaré knew similar moments of deep intuition that arrived unbidden, "a manifest sign", he thought, "of long, unconscious prior work".

Holton writes with relish of a conversation on the origins of the uncertainty principle between Heisenberg and Einstein in the mid-1920s that Heisenberg recounted to him in 1956. But Holton finds Heisenberg's politics appalling, and rebukes him for his willingness to collaborate with the Nazi regime and for issuing "astonishing exaggerations" about Einstein's role in the atomic-bomb project while claiming that he had declined on moral grounds to build an atomic bomb for Hitler.

Holton rightly insists that the Heisenberg in Michael Frayn's play *Copenhagen*, who said he knew how to build a bomb but refrained, is a fictional character and ought to be viewed as such.

Holton is dismissive of the postmodern critique of science, saying it holds that the aim of achieving objective truth is unrealizable "because there is no difference between the laws scientists find in nature and the arbitrary rules that govern baseball games". He finds part of its roots in nineteenth-century European romanticism, which was at times scientifically productive. But he also sees shades of it in Hitler's declaration that "there is no truth, in either the moral or the scientific sense". For Holton, truth emphatically exists in both senses. It is clear from these graceful essays that he stands with Rabi, admiring his insistence that science is an essential part of culture, an ennobling activity, a guide to objective thinking and a "unifying force for all of humanity".

Daniel J. Kevles is in the Department of History, Yale University, New Haven, Connecticut 06520-8324, USA.

Part of the problem for science has been attempting to distil a working definition of genius that removes its more subjective and untestable historical and cultural associations, while still retaining our idea of it. This is far from easy. One tenet is that a genius must be recognized as such by the relevant experts in the field — but by that reckoning, if Einstein hadn't published his theories, he would have been barred from the title. Despite the many difficulties with investigating genius (hence the mixed results), science has tried to break it down into components such as intelligence, structure and function of the brain, madness, level of disinhibition, even genetic inheritance.

Because of the somewhat elusive definition of creativity, Nancy Andreasen opts for a case-study approach in her book *The Creating Brain*. Andreasen is an MD with a PhD in Renaissance English literature, which formed the basis for her first book, *John Donne* (Princeton University Press, 1967). From Mozart to August Kekulé, and Henri Poincaré to Samuel Taylor Coleridge, she unravels the insights, accounts and descriptions of their moments of revelation. After dissecting their multifarious personality traits, she attributes their extraordinary creativity in part to "brains that are more facile at creating free associations", and to contributions from the "unconscious mind". Her accounts suggest that unconscious processes are at work, but as the US writer Gertrude Stein warned us, they cannot be summoned at will: "It takes a lot of time to be a genius, you have to sit around so much doing nothing, really doing nothing." Perhaps that's some comfort for us mere mortals.

No account of creativity would be complete without a departure into the notion that genius and mental illness are inextricably linked. There is a pervasive belief that creativity and bipolar disorder, in particular, have a strong connection — perhaps we like to think that in order to be creative one must, at the very least, have a touch of madness. Andreasen recounts her own experience investigating individuals from the Iowa Writers' Workshop, who to her surprise had an increased incidence of depression, either bipolar or unipolar, suggesting a "relationship between artistic creativity and mood disorders". It is interesting to speculate whether this relationship is causal, is specific to certain subpopulations of mental illness, or whether the arts provide a suitable home for those with a particular illness. Whatever the reason, the link is compelling, and it is easy to produce a list of names that provide anecdotal support. But why do so few of those who are debilitated by bipolar disorder receive the benefits of this extraordinary artistic creativity?

It is well recognized that brain development occurs on a hectic timetable, given that several trillion synaptic connections must be laid down for the brain to function at average levels. During early pregnancy, 250,000 brain cells are created every minute, and this continues at a ferocious rate during infancy, when

The making of a genius

The Creating Brain: The Neuroscience of Genius

by Nancy C. Andreasen

Dana Press: 2005. 225 pp. \$23.95

Mark Lythgoe

Ever since the first bright spark discovered how to make fire, the recipe for genius has been one of culture's most alluring quests. Yet historically, our conception of genius has been

mysterious. The very idea that it could be explained seems to run counter to its essence. From antiquity until the Enlightenment and beyond, genius was seen as an innate trait bestowed by the gods. But as the gods lost their power, it has fallen to others to do the explaining. Even modern science has been reluctant to take up the challenge, as the apparent unpredictability of creative genius seems to elude any singular systemic explanation.



connections form that allow you to crawl, walk and then talk. There is then a process of constant organization and reorganization that continues until early adult life and beyond. This process, known as brain plasticity, is the basis for Andreasen's self-help guide to improved creativity. She advises us to perform mental exercises, explore unfamiliar fields of endeavour, meditate or "just think", practise observing, describing and imagining. And kids

must turn off the TV, read, explore the natural world and listen to classical music. Despite what might seem like reasonable offerings, this section, and maybe others too, could perhaps have been complemented by a reference list to allow some assessment of the arguments and suggestions presented.

Andreasen writes with clarity and ease, interspersing personal and scientific opinion. She makes wonderful connections between

the arts and sciences, which surely spring from her background in literature. And she provides a succinct overview of diverse fields of investigation, as well as providing a perspective that reaches beyond the usual approaches to understanding the relationship between creativity and the brain. ■

Mark Lythgoe is a neurophysiologist in the Radiology and Physics Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.

Digging for clues

Discovering Dorothea: The Life of the Pioneering Fossil-Hunter Dorothea Bate

by Carolyn Shindler

HarperCollins: 2005. 304 pp. £25

Jennifer Clack

The slightly blurred, somewhat ghostly figure caught in a pose of resolute determination on the cover of this book is a highly appropriate image and captures the essence of the subject. How do you write a biography of someone who left virtually no personal documents, but a wealth of published scientific articles? The author, Carolyn Shindler, faced this problem when she tackled the life of Dorothea Bate, a pioneering female palaeontologist who worked in the first half of the twentieth century.

Bate's interest in natural history and fossils began early, when she was about ten. It seems to have arisen spontaneously, rather than from the influence of any adult around her — this is often the way with palaeontologists and natural historians. The absence of any personal diaries, from any stage of her life, leaves unanswered questions, such as what motivated her initially, and what drove her to continue against a multitude of difficulties. Her initiative in beginning such daunting adventures as expeditions to remote and poorly resourced locations with only sporadic, sometimes unreliable, local support was exceptional, and leaves me feeling inadequate. What's more, it is clear that she had to face parental opposition and relative poverty from time to time. On the other hand, she received appreciative support from professional palaeontologists at the British Museum (Natural History) in London — male, of course — who recognized her unique contributions. If I have a criticism of Shindler's writing, it is that, in the early parts of the story, the difficulties are somewhat overemphasized to become almost tedious, whereas the successes are downplayed.

She was remarkable for more than being a female palaeontologist at a time when the discipline, and its locations, were male dominated. She also pioneered collecting from previously unexplored and almost inaccessible parts of several Mediterranean islands, discovering new species and faunas from the Pleistocene of the area, and demonstrating



the idea, novel at the time, that island dwarfing of elephants and hippos occurred in parallel on several islands.

As Bate grew in experience and academic stature, she began to integrate evidence from many sites and faunas to infer climatic changes over recent millennia, at a time when such thinking was in its infancy. She later incorporated this evidence with new finds in archaeology and anthropology to place human remains in their faunal context. She was among the first to recognize that the animals associated with ancient human habitations could shed considerable light on human activities and ecology, and she brought ideas of climate change to bear on human evolution through the Pleistocene. I was previously unaware of her work (my work deals with Palaeozoic fossils), but my colleagues who work on Pleistocene or Quaternary material not only know of her but continue to use the

material she collected. Her ideas and techniques were ahead of her time. Her extensive publication record began in 1901 and continued to grow in depth and understanding, and with undiminished energy, until 1955. Despite this, it was not until near the end of her life that she gained permanent paid full-time employment, at the British Museum (Natural History)'s site in Tring.

In her later endeavours, she developed friends and colleagues in the archaeological world, several of whom were also women, and some went on to be pioneers in other respects. Dorothy Garrod, for example, was the first woman professor at the University of Cambridge. It is as though archaeology was already seen as a field in which women could play a significant role.

This biography could perhaps be criticized for its lack of in-depth analysis of the subject's personality or psyche, or that bringing to

attention a 'forgotten woman in a man's world' is passé in the early twenty-first century. However, I think the author is justified on other grounds. With almost all of Bate's personal records, such as diaries or photographs, having been lost, destroyed or perhaps never made, Shindler has focused instead on the work of a scientist of considerable ability, originality and resourcefulness. Perhaps it is today's almost voyeuristic obsession with the analysis of motives and feelings, and the cult of personality, that leads us to expect a biography to address such issues — especially as the

subject is female, and therefore 'ought' to have recorded her innermost feelings. I certainly don't keep any such diaries, and I wonder how many of my fellow scientists do.

In this biography, then, we are dealing first and foremost with a scientific record. In the end, it is the results of her research — the new specimens, species and their descriptions — that stand the test of time. I am grateful to Shindler for bringing Bate and her work to my attention. ■

Jennifer Clack is at the University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, UK.

A Stone Age greenhouse

Plows, Plagues and Petroleum: How Humans Took Control of Climate

by William F. Ruddiman

Princeton University Press: 2005. 272 pp.

\$24.95, £15.95

Robert J. Charlson

The activities of Stone Age farmers may have altered Earth's climate. This is the exciting but controversial theory conveyed by palaeoclimatologist William Ruddiman in his well-written book *Plows, Plagues and Petroleum*.

I am not a climatologist, but my work on atmospheric chemistry, aerosols and cloud physics relates to Ruddiman's analysis of Earth's climate over the past few millennia. Aerosols and clouds must be included in any analysis of palaeoclimate because they are so variable and exert such a powerful influence on the albedo of the planet. It cannot be assumed that they remained constant over this time. I strongly support Ruddiman's view that fitting all these pieces together to figure out the key cause-and-effect relationships "makes studying climate history fun".

Ruddiman's book is unusual because he candidly describes his main ideas as a thesis rather than as fact, and states that they are currently being debated in refereed publications. This high level of candour will certainly be appreciated by scientists, and the book's descriptive analogies and lack of jargon make it accessible to the lay reader as well.

The book starts with the importance of climate to human history, the basic science of climate, the connections between Earth's orbit and climate, and the modulation of ice ages, monsoon circulations and climate, with "nature in control".

The main thesis then emerges, with a description of the inexorable changes imposed by humans on the chemical composition of Earth's atmosphere. Starting with Ruddiman's strongest scientific argument, the increase of the greenhouse gas methane about 5,000 years ago is attributed to Stone Age farming. Simple calculations of the amount of methane generated per person by flood irrigation, animal

husbandry and biomass burning show emission rates high enough to explain the methane data obtained from ice cores.

Other population-based estimates for the emission of carbon dioxide, another greenhouse gas, show that over a period of several thousand years, enough biomass combustion could have occurred during land clearing to increase the concentration by the observed amount, 40 parts per million. Stone Age humans apparently began burning forests about 8,000 years ago, which fits with the time series of carbon dioxide data.

Moving from the data-based analysis of rising levels of methane and carbon dioxide to the tentative business of climate forecasting, Ruddiman then asks: "Have we delayed a glaciation?" He thinks we have, arguing that the stability of the climate over the past 10,000 years "may have been an accident". The warmth of the past several thousand years "stems from a colossal coincidence: a natural cooling" that was "offset by a human-induced warming".

Then, in what almost seems to be an afterthought, the small 'wiggles' in the carbon dioxide record and a decrease in the rate of its increase over the past 2,000 years are interpreted to have been caused by the decimation

of human populations in epidemics and pandemics. Smaller populations produced less carbon dioxide, adding yet more variables to the study of palaeoclimate.

The main part of the book concludes with three chapters on the nature of climate in a future with "humans in control". Ruddiman provides a vivid description of the immensity of human influence on the climate and on the environment in general. The growth of greenhouse gases is singled out, with carbon dioxide and methane being dominant. The paradoxically small observed temperature increase in the industrial period (about 0.6 °C), compared with the larger rise of 0.8 °C caused by previous human activity, is attributed to a delay in climate response and to cooling factors generated by humans, such as sulphate aerosols.

Current global warming is then analysed over a shorter time scale (centuries) by means of uncertain estimates of the future rate of carbon dioxide and methane emission by human activities. This shorter-term analysis is then set in the context of the distant past (with nature in control) and for several centuries in the future (under human influence), all properly labelled as uncertain.

An epilogue completes the book, in which Ruddiman presents his own views on the issues of climate change. He decries the habit in the press of wanting "clever, crisply phrased sound bites" and he laments the polarization regarding global warming. He calls for recognition of the underlying reality that "draconian economic sacrifices" will be needed, and that this should put the global-warming debate "in a clearer perspective". So ends an excellent book summarizing and placing in context the age-old influence of humans on atmospheric composition, climate and global warming.

However, just as it is difficult to prove that Earth's warming over the industrial period has been caused by the emission of greenhouse gases through human activities, it is also difficult to prove that Stone Age farming caused a lack of cooling thousands of years ago. The debates about the causality of both present-day warming and the warming needed in the past to delay an ice age can be summarized briefly in terms of the concept of climate forcing, the imposed change in energy balance. For causality of global warming to be demonstrated, three interconnected premises must hold: the anthropogenic global forcing must be positive and substantial in magnitude; the temperature change must be positive and beyond the range of natural variability; and the former must be the cause of the latter.

If Ruddiman used the concept of forcing to strengthen and refine his thesis, he could address more specifically the uncertainties in assigning the causality of temperature increases to Stone Age agronomists. He could find further reasons for the higher sensitivity of temperature to early changes in methane and carbon dioxide levels. He would have to consider smoke from biomass burning as a



cooling factor — might it have played a role in causing the little ice age?

The debate continues among climatologists, but it would seem that Ruddiman must be at least partly correct: there can be no question that Stone Age farmers added methane and carbon dioxide to the atmosphere, but how much and when? The debate, which seems to centre on the interpretation of details of

temperature inferences from isotopic records in sediments and ice cores, cannot disguise the fact that levels of both methane and carbon dioxide began to increase long before the onset of industrialization. ■

Robert J. Charlson is in the Department of Atmospheric Sciences and the Department of Chemistry, University of Washington, Seattle, Washington 98125, USA.

Sticking with nature

The Gecko's Foot. Bio-inspiration: Engineered from Nature.

by Peter Forbes

Fourth Estate: 2005. 272 pp. £20

R. McNeill Alexander

Biomimetics is the application of ideas from nature in engineering. At first sight, it seems a promising approach: evolution by natural selection is extremely effective, and designers can surely learn from its solutions. But so far, biomimetics has not achieved very much. As recently as 2003, Julian Vincent, professor of biomimetics at the University of Bath, UK, and one of its keenest proponents, wrote that “the only successful examples I know of are Velcro (1955) and the Anglepoise lamp”.

I would discount the Anglepoise lamp. Its springs will support it in any position, much as the elastic ligament in a cow's neck supports its head, but I know of no evidence that the lamp designer knew about the cow. Velcro is more convincing. Its inventor, George de Mestral, got the idea from burs that stuck to his dog's

fur on a hunting expedition. If he had copied the bur-fur attachment exactly, he would have put hooks on only one of the surfaces. Instead, he put hooks on both. Biomimetics does not imply slavish copying, merely the application of a principle from nature. For this reason, Peter Forbes prefers the term ‘bio-inspiration’. His new book about it, *The Gecko's Foot*, is designed for general readers, who will need very little prior scientific knowledge.

Surprisingly, Forbes starts his book with a discussion of scale. He argues that there is a range of sizes to which we are relatively blind. We can see things that are as small as a micrometre with optical microscopes, and chemistry has taught us a great deal about atoms and molecules of the order of a nanometre. He argues that the gap between those limits is our blind zone, and is especially promising for biomimetics.

Forbes supports this argument with an excellent example, a recently developed paint inspired by lotus leaves, which stay clean even in the muddiest water. Wilhelm Barthlott, a

botanist at the University of Bonn, Germany, discovered that these leaves are covered in tiny hydrophobic tubercles, in the blind-zone range of sizes. On this surface, water forms almost spherical drops, and rolls off as if from a duck's back, carrying dirt away with it. (The duck's back works on the same principle.) Barthlott persuaded a paint company that the lotus effect, as he called it, had commercial potential, and they developed a paint for the outside of buildings. It looks like ordinary paint but has a silicone surface that is rough in the blind-zone scale. Rain keeps it beautifully clean.

Forbes continues with other examples. The gecko's foot, which gives his book its title, is covered with a carpet-like pile of hairs. The end of each hair has many branches a few hundred nanometres across. These hairs make such close contact with the surfaces on which the lizard walks that they adhere by intermolecular attraction (the van der Waals effect). This lets the gecko climb walls and even run across the ceiling. Adhesive tape working on the gecko principle has been made but not yet commercially developed.

Forbes discusses other examples in the size range that he thinks so promising, including the multilayer reflectors in the wings of iridescent butterflies and the self-assembling properties of bacteriophages but, as with the gecko, these are potential rather than realized commercial applications. He discusses some larger-scale effects, including the aerodynamics of insect flight and the folding pattern that enables a large leaf to be opened from a slender bud, but these are also of possible future value rather than current successes.

Forbes frankly admits that biomimetics has had few triumphs so far. He tells us in his final chapter: “I have written about science and technology on the hoof because this is a new science and success is a matter of decades of work. Rather than wait until some of these technologies have become commonplace, I have tried to capture the Wordsworthian ‘bliss-it-was-in-that-dawn-to-be-alive’ moment of seeing what was, until 15 years ago, a wholly unexpected science taking shape before our eyes.”

There is a great deal of interesting and unfamiliar material here, but I found some of the scientific explanations unsatisfactory. To quote one example, in the explanation of the lotus effect, the statement that “water sits on the points of the bumps, with the compression of the air in the cavities giving extra buoyancy” is misleading: the effect depends on contact angles, not air pressure and buoyancy. On the other hand, an admirably comprehensive set of notes directs readers to the primary scientific literature. The general readers for whom this book is intended will enjoy it, and it will give them some appreciation of a fascinating field. ■

R. McNeill Alexander is in the Faculty of Biological Sciences, Miall Building, University of Leeds, Leeds LS2 9JT, UK.



NEUROBIOLOGY

Triggers for channel opening

Cynthia Czajkowski

Fast transmission between nerve cells relies on specialized ion channels. Probing the structure of these proteins reveals how the binding of a neurotransmitter causes the communication channels to open.

Chemical signalling in the brain involves the rapid opening and closing of channels known as ligand-gated ion channels, which lie in the membranes of nerve cells. Binding of a specific activator (a ligand) to these proteins triggers the opening of an integral pore through the membrane in as little as tens of microseconds¹. Although we know a fair amount about the structure of ligand-gated ion channels, the mechanisms by which the binding of a ligand triggers channel opening are still under debate. On page 243 of this issue, Lee and Sine² identify a network of interacting amino-acid residues in one such protein, and reveal a pathway by which changes at the protein's ligand-binding site can be propagated to its channel region. And on page 248 Lummis and colleagues³ identify a proline residue that acts as a molecular switch to control channel opening. Together, the two reports provide a compelling description of the structural machinery that couples ligand binding to channel gating.

Communication between nerve cells takes place at junctions called synapses. When a presynaptic cell is activated, it releases neurochemicals (neurotransmitters) across the synapse that bind to ligand-gated ion channels on the surface of the postsynaptic cell. Binding of neurotransmitter causes the channels to open, allowing ions to flood across the postsynaptic-cell membrane and change the cell's activity. So ligand-gated ion channels can be thought of as transducers that rapidly convert chemical signals into an electrical output. Their opening and closing regulate information flow throughout the brain, and mutations in these channels are responsible for a number of 'channelopathies', such as congenital myasthenic syndromes, epileptic disorders and hereditary hyperekplexia.

Lee and Sine² and Lummis *et al.*³ examined the structures of two members of the 'Cys-loop' family of ligand-gated ion channels. This family includes channels that respond to the neurotransmitters acetylcholine, serotonin, γ -aminobutyric acid (GABA) and glycine. The receptors are large transmembrane proteins (molecular weight 300,000) consisting of five similar subunits arranged around a central ion-conducting channel, with

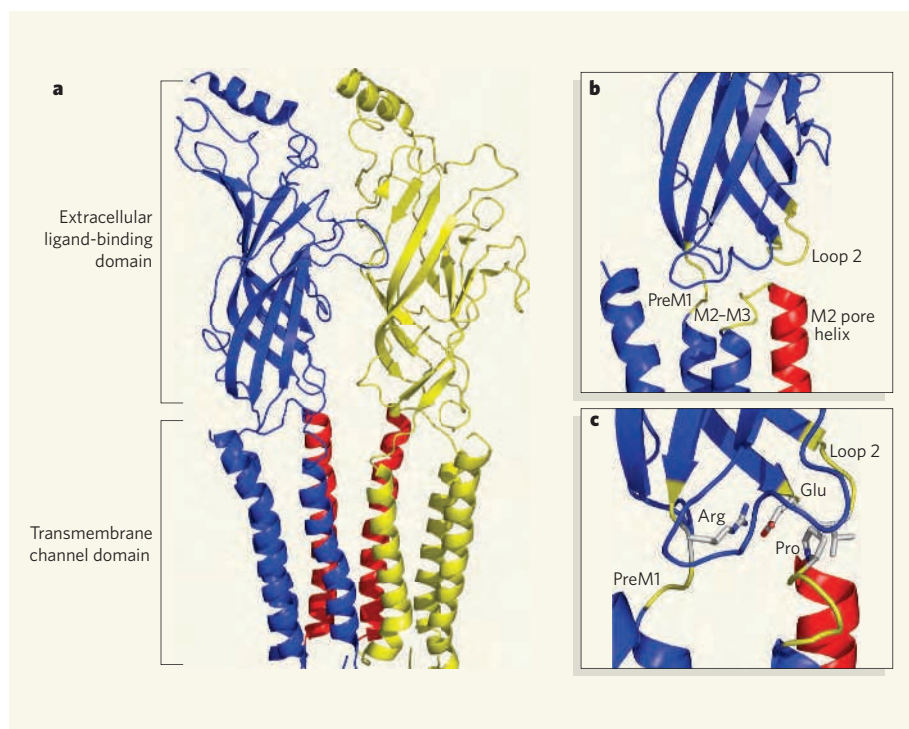


Figure 1 | The pathway that links ligand binding to gating. **a**, Structure of two subunits of the nicotinic acetylcholine receptor. The M2 helices that line the channel pore are shown in red. **b**, **c**, Close-up of the interface between the extracellular ligand-binding domain and the intracellular transmembrane channel domain of one subunit. **b**, Lee and Sine² identified several regions — preM1, loop 2, M2–M3 (shown in yellow) — that are interconnected and couple ligand binding to channel opening. The pore-lining helix (M2) is red. **c**, Some of the specific amino-acid residues pinpointed by Lee and Sine that form the activation pathway involved in this coupling (Arg, Glu, Pro). Lummis *et al.*³ found that the proline residue at the same position in the serotonin 5-HT₃ receptor acts as a switch to control the ion-channel gating.

each subunit contributing to the lining of the transmembrane channel (Fig. 1a). The neurotransmitter binds to the extracellular interface between two subunits. But what has long puzzled researchers is how the binding of a neurotransmitter, which is around 6 Å long, is translated so rapidly into the opening of an ion channel more than 50 Å away in the transmembrane domain of the receptor.

Lee and Sine² set out to answer this question. They used the nicotinic acetylcholine receptor, whose structure was recently refined to 4-Å resolution⁴, to identify receptor amino acids that could physically link the binding site to the channel. They then created

a series of mutations, by substituting amino acids, to break these potential links, and analysed the mutations' effects, both individually and in combinations, on channel activity. As a result, they identified a set of interacting residues that functionally and structurally link the binding site to the channel.

The interacting residues connect three distinct regions of the receptor — the preM1 region, loop 2 and the M2–M3 loop. These regions lie at the interface between the binding site and the channel, and so are perfectly situated to transmit changes in the binding site to the channel (Fig. 1b). Although previous studies^{5–7} have pinpointed these regions as

coupling elements and have even identified some pair-wise interactions, the real novelty of Lee and Sine's work is that it teases apart the functional interactions between these different regions on an atomic scale. Their data provide a detailed molecular scaffold for the idea that the binding of neurotransmitter triggers a wave of conformational changes that propagates from the ligand-binding site to the pore through the membrane⁸.

One particularly attractive feature of Lee and Sine's model is the electrostatic interaction that exists between a positively charged arginine residue in the preM1 region and a negatively charged glutamate residue in loop 2 (Fig. 1c). Charged residues occur in these positions in every member of this receptor superfamily, suggesting that it is a common mechanism for linking binding-site changes to loop 2. Loop 2 sits above the extracellular end of the M2 helix that forms the channel, next to the M2–M3 loop, and one can easily envisage how movements in loop 2 could be conveyed to the channel through interactions between residues in these two regions^{9–11}. Lee and Sine found several interacting residues, in particular a proline in the M2–M3 region (Fig. 1c). And this is where the work by Lummis *et al.*³ comes in — their study concentrated on a proline in exactly the same position in the serotonin 5-HT₃ receptor.

Proline residues are unique in that their side chains are covalently bonded to the nitrogen atom of the protein's peptide backbone. The nitrogen atom of a peptide bond is usually involved in secondary interactions with other parts of the protein, helping to hold helices together, for example; but at proline, these secondary interactions are disrupted. In addition, proline is the only natural amino acid for which two different conformations of the peptide bond (*cis* and *trans*) are possible. These properties allow prolines to act as molecular hinges or switches¹².

In an elegant study on 5-HT₃ receptors, Lummis *et al.*³ replaced the proline in the M2–M3 loop with a series of synthetic amino acids that had different propensities for adopting *cis* versus *trans* conformations. Amino acids that favoured the *cis* conformation resulted in receptors with high apparent affinities for neurotransmitter, producing ligand-induced 'locked' open channels. Amino acids that mostly took on *trans* conformations resulted in unresponsive closed channels. So it would seem that the conformation of the M2–M3 proline is coupled with the conducting state of the channel, pointing to the proline as a gating switch.

Taken with Lee and Sine's work, this conjures up a mechanism in which neurotransmitter binding to the receptor sets in motion a cascade of structural movements that end up flipping the proline to the *cis* conformation. This results in a repositioning of the M2 channel-lining region, such that the channel pore opens.

Although the two studies tell a compelling story of how ligand binding is translated into channel opening, it is probably not complete. The proline residue in the M2–M3 loop of the acetylcholine and serotonin receptors does not occur in the receptors for GABA or glycine, so other mechanisms must be invoked to explain how they open. Changes between *cis* and *trans* proline conformations are generally slow, so we need to determine whether this switch is fast enough to open channels on a sub-millisecond timescale. And because other regions and residues are involved in coupling binding to gating, the activation pathway charted by Lee and Sine will probably have others feeding into it. Finally, ligand-gated ion channels not only open and close in response to binding neurotransmitter, but they also become desensitized (close) in the continued presence of neurotransmitter. How this occurs is still a mystery. ■

MATERIALS SCIENCE

Erasing electron mass

Charles L. Kane

Two-dimensional graphite could be useful in carbon-based electronic devices. How electrons move in these structures seems best described by relativistic quantum physics, modelling them as if they have no mass at all.

Graphite, the form of carbon found in pencil lead, leaves its mark thanks to weakly coupled layers of atoms that slide easily over one another. A single such layer — a two-dimensional sheet of carbon a single atom thick — is known as graphene. Although graphene has been studied for decades, graphene was only isolated in 2004 after a long struggle. The successful method was astonishingly simple: starting with a graphite crystal, layers of carbon atoms were peeled off one by one with adhesive tape, until a single-layer flake was left¹. In this issue, Novoselov *et al.* (page 197)² and Zhang *et al.* (page 201)³ investigate the properties of graphene further, showing it to be a remarkable conductor in which electrons mimic the behaviour of massless, relativistic particles.

In graphene, the carbon atoms are arranged in a honeycomb pattern (Fig. 1a), with each atom bound to three neighbours through strong, covalent bonds. This gives graphene exceptional structural rigidity within its layers. Because a carbon atom has four electrons available for bonding, each atom also contributes one unbound electron that is free to wander through the crystal, giving graphene its second distinctive characteristic — excellent conductivity.

The mobile electrons in graphene seem to behave differently, however, from those in two-dimensional semiconductor structures. In semiconductors, an electron can be

Cynthia Czajkowski is in the Department of Physiology, University of Wisconsin-Madison, 601 Science Drive, Madison, Wisconsin 53711, USA. e-mail: czajkowski@physiology.wisc.edu

1. Chakrapani, S. & Auerbach, A. *Proc. Natl Acad. Sci. USA* **102**, 87–92 (2004).
2. Lee, W. Y. & Sine, S. M. *Nature* **438**, 243–247 (2005).
3. Lummis, S. C. R. *et al.* *Nature* **438**, 248–252 (2005).
4. Unwin, N. J. *Mol. Biol.* **346**, 967–989 (2005).
5. Hu, X. Q., Zhang, L., Stewart, R. R. & Weight, R. R. *Nature* **421**, 272–275 (2003).
6. Kash, T. L. *et al.* *Nature* **421**, 272–275 (2004).
7. Kash, T. L., Dizon, M. J., Trudell, J. R. & Harrison, N. L. *J. Biol. Chem.* **279**, 4887–4893 (2004).
8. Chakrapani, S., Bailey, T. D. & Auerbach, A. *J. Gen. Physiol.* **123**, 341–356 (2004).
9. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. *Nature* **423**, 949–955 (2003).
10. Lester, H. A., Dibas, M. I., Dahan, D. S., Leite, J. F. & Dougherty, D. A. *Trends Neurosci.* **27**, 329–336 (2004).
11. Reeves, D. C., Jansen, M., Bali, M., Lemster, T. & Akabas, M. J. *Neurosci.* **25**, 9358–9366 (2005).
12. Samson, M. S. P. & Weinstein, H. *Trends Pharmacol. Sci.* **21**, 445–451 (2000).

modelled as a particle that obeys Newton's laws of motion, provided it is ascribed an effective mass, m^* , which takes into account the interaction between the electron and the semiconductor's crystal lattice. Electrons in semiconductors are thus characterized by a quadratic relationship between energy and momentum (Fig. 1b), and their quantum-mechanical behaviour can be described by the non-relativistic quantum theory formulated in the Schrödinger equation.

In contrast, the observations of Novoselov *et al.*² and Zhang *et al.*³ show that, in the honeycomb structure of graphene, the relation between energy and momentum of the conduction electrons is linear (Fig. 1c). This is reminiscent of Einstein's theory of relativity for massless particles — which travel at the speed of light — and suggests that electrons in graphene obey a two-dimensional version of the relativistic quantum theory introduced by Paul Dirac in 1928. Electrons in graphene are, of course, not actually massless, and their typical speed ($8 \times 10^5 \text{ m s}^{-1}$) is almost 400 times lower than the speed of light in a vacuum (about $3 \times 10^8 \text{ m s}^{-1}$) — although still much higher than the speed typical of electrons in semiconductors.

The conclusions of the authors^{2,3} are based, to a large extent, on their observation of the quantum Hall effect in graphene. This phenomenon is the quantum-mechanical

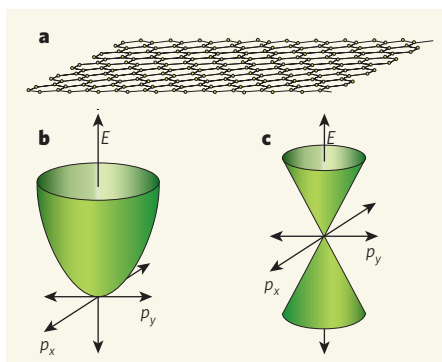


Figure 1 | Slim, strong and a live wire. **a**, The honeycomb lattice pattern of graphene explains its strength and good conductivity. Each carbon atom (green dot) uses three of its outer valence electrons to form strong covalent bonds, leaving one left over that is available for conduction. **b**, The quadratic, newtonian energy-momentum relation, $E = p^2/2m^*$ (E , energy; p , momentum; m^* , reduced mass) is obeyed by electrons in a semiconductor. **c**, The energy-momentum relation of electrons in graphene is quite different, $E = v|p|$ (v is the electron velocity), allowing them to be modelled as massless, relativistic particles according to the Dirac formulation of quantum mechanics.

analogue of an effect observed by Edwin H. Hall in 1879 in a macroscopic conductor. In a magnetic field, charged particles such as electrons experience a force perpendicular to their motion, causing them to move in closed circles. A magnetic field applied at right angles to the surface of a conductor will therefore deflect the electrons from their path between the terminals, establishing a voltage perpendicular to the direction of current flow. In a conventional solid, the Hall resistance (given by the ratio between the perpendicular voltage and the forwards electron current) increases smoothly as the applied magnetic field increases.

In a two-dimensional solid at low temperatures, however, quantum effects — specifically, the wave-like properties of the electrons — come into play. Only an integer number of electron wavelengths may fit into the circular orbits induced by the magnetic field, restricting the permitted electron energies to a set of discrete, quantized values. This means that in turn the Hall resistance no longer increases continuously with increasing magnetic field strength, but in a characteristic series of steps quantized in units of h/e^2 (where h is Planck's constant and e the electron's charge). Since its discovery in 1980, this quantum Hall effect has had a profound impact on semiconductor research, with two Nobel Prizes in Physics — those of 1985 and 1998 — being awarded for work on it.

Both Novoselov *et al.*² and Zhang *et al.*³ observed quantized steps in the Hall resistance of graphene — a result that highlights the exemplary quality of their samples. The exact numerical values at which the resistance steps occurred were, however, found to be shifted by one-half of a unit from those expected for non-relativistic electrons. The relativistic

theory provides an explanation: just as Dirac's original theory explained why electrons have an intrinsic angular momentum, known as spin, the effective Dirac theory for graphene endows electrons with an additional 'pseudospin'. When an electron completes a circle in an applied magnetic field, its pseudospin rotates by 360° . As is the case for real spin, such a rotation introduces a 180° phase shift in the electron wave, so an additional half wavelength must fit in the circumference of the circle, changing the pattern of allowed energies. The pattern of the observed steps^{2,3} in the Hall resistance fits with this picture perfectly, providing convincing evidence for graphene's Dirac electronic structure.

The electrical conductivity of graphene at zero magnetic field, with a minimum value at low temperature that is close to $4e^2/h$ for several different samples², raises interesting

questions. Such a 'universal' minimum conductivity is in itself reasonable; how to explain the precise value, however, is an open question. Away from the minimum, the conductivity varies linearly with electron density — expected for newtonian electrons, but surprising for relativistic electrons. Finding answers to these remaining riddles will be essential for realizing the potential of graphene-based electronics. So it's time for theorists to sharpen their pencils.

Charles L. Kane is in the Department of Physics and Astronomy, University of Pennsylvania, 209 S. 33rd Street, Philadelphia, Pennsylvania 19104, USA.

e-mail: kane@physics.upenn.edu

1. Novoselov, K. S. *et al.* *Science* **306**, 666–669 (2004).
2. Novoselov, K. S. *et al.* *Nature* **438**, 197–200 (2005).
3. Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. *Nature* **438**, 201–204 (2005).

MICROBIOLOGY

RAMP resistance

Angus Buckling and Michael Brockhurst

There is an urgent need for new antimicrobial agents because antibiotic resistance has become so prevalent. But a promising class of such agents, known as RAMPs, may suffer from the same problem.

In a report published in *Proceedings of the Royal Society*, Perron *et al.*¹ demonstrate experimentally that bacteria can readily evolve resistance to a group of proteins called ribosomally encoded antimicrobial peptides (RAMPs). RAMPs are produced by animals, plants, fungi and bacteria as part of their natural defence against microbial attack^{2–4}, and are being developed as antibiotics. But because bacterial resistance to chemotherapeutic RAMPs could confer resistance to the battery of innate human RAMPs⁴, the worrying prospect is that widespread use of these agents may compromise our natural defence against bacteria.

With the emergence of bacteria that are resistant to 'last resort' antibiotics such as vancomycin⁵, there is a desperate need to identify new antimicrobial agents. RAMPs may be just such agents. They are a diverse group of proteins, and their mode of action varies considerably, but a common feature is their positive charge. This allows them to bind to the negatively charged membranes of bacteria. The effectiveness of a variety of RAMPs in clinical trials^{2,3}, and the recent discovery of fungus-derived RAMPs that can be produced in large yields³, suggest RAMPs could be in widespread clinical use within the next few years.

The potential advantages of RAMPs are the apparent difficulty that bacteria face in evolving resistance to them, and the fact that resistance to conventional antibiotics does not seem to confer resistance to RAMPs². Bacteria have

obviously encountered RAMPs in one form or another for millions of years, yet widespread resistance is rare. Furthermore, previous experimental tests suggest that the evolution of resistance does not readily occur^{6,7}.

However, resistance evolution is all about the level of exposure. Although bacteria had been exposed to natural antibiotics such as penicillin and streptomycin (respectively produced by the *Penicillium* mould and *Streptomyces* bacteria) for millions of years, resistance was at low levels when widespread clinical use of these drugs began in the 1940s. But after a few years of exposure to high clinical doses, resistance was widespread in many species of pathogenic bacteria.

On the basis of such logic, Perron *et al.*¹ attempted to experimentally induce resistance to a RAMP in two different bacterial species, *Escherichia coli* and *Pseudomonas fluorescens*. The RAMP in question, pexiganan, is a synthetic analogue of a RAMP derived from toads (magainin) that has been modified for use as a chemotherapeutic agent. The authors exposed bacteria to slowly increasing concentrations of the drug for 600 generations (a few months in the lab), unlike previous work where drug concentrations were kept constant, and populations were allowed to evolve for no more than 200 generations^{6,7}. The results were astounding: 22 out of 24 populations of bacteria had developed resistance to the drug.

The ability of bacteria to evolve resistance to

antibiotics does not necessarily mean that resistance will become a widespread problem. Antibiotic resistance often compromises the bacteria in other ways — for example, by reducing their growth rate⁸. This means that antibiotic-sensitive bacteria will outcompete the resistant forms when neither is exposed to the antibiotic. Perron *et al.*¹ investigated this possibility, and indeed found a 'cost' of antibiotic resistance: in the absence of the antibiotic, resistant bacteria took longer to start reproducing than control bacteria, although once they had got going, their replication rate was unaffected.

Unfortunately, bacteria have other tricks up their sleeves. In addition to adapting to antibiotics, they can also adapt to antibiotic resistance. There have been numerous cases of bacteria with antibiotic resistance developing mutations in other parts of their genome that compensate for the associated costs^{8,9}. These adaptations are sometimes so specific that the growth rate of bacteria can decrease if the genetic changes conferring antibiotic resistance are replaced with the original sensitive form of the gene after compensatory adaptation has occurred⁹.

Why should bacterial resistance to RAMPs cause more concern than resistance to other antibiotics? The major problem will be if resistance to chemotherapeutic RAMPs also confers resistance to naturally occurring RAMPs in humans and other organisms. Bacteria that are normally dealt with unnoticed by our innate immune system may then cause serious infections. Large-scale use of chemotherapeutic RAMPs may ultimately help pathogenic bacteria colonize parts of animals and plants that were previously off limits to them.

This perspective may be overstating the case for concern. Humans alone produce a highly diverse arsenal of RAMPs, which are also thought to be constantly evolving new ways of targeting bacteria¹; and RAMPs constitute only one part of our natural immunity. Furthermore, RAMP resistance, where observed, is often specific to a small range of RAMPs⁴. There are exceptions, however. A variety of bacteria, including *Staphylococcus aureus* — famed for methicillin resistance — and the opportunistic pathogen *Pseudomonas aeruginosa* have evolved a degree of generalized RAMP resistance by increasing the amount of positively charged protein in their membranes. The consequence may be to reduce the binding efficiency of the positively charged RAMPs¹⁰.

As Perron *et al.*¹, and others²⁻⁴, emphasize, RAMPs are likely to make a major contribution to human health and agriculture. But given the prospect of resistance, extra caution is necessary in developing and using them. ■ Angus Buckling is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. Michael Brockhurst is at ISEM, Université de Montpellier II, 34095 Montpellier, France. e-mail: angus.buckling@zoology.oxford.ac.uk

1. Perron, G. G., Zasloff, M. & Bell, G. *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2005.3301 (2005).
2. Zasloff, M. *Nature* **415**, 389–395 (2002).
3. Mygind, P. H. *et al.* *Nature* **437**, 975–980 (2005).
4. Bell, G. & Gouyon, P.-H. *Microbiology* **149**, 1367–1375 (2003).
5. Trevor, F. C. & McDonald, L. C. *Curr. Opin. Infect. Dis.* **18**, 300–305 (2005).

6. Ge, Y. *et al.* *Antimicrob. Agents Chemother.* **43**, 782–788 (1999).
7. Steinberg, D. A. *et al.* *Antimicrob. Agents Chemother.* **41**, 1738–1742 (1997).
8. Maisnier-Patin, S. & Andersson, D. S. *Res. Microbiol.* **150**, 360–369 (2004).
9. Schrag, S. J. *et al.* *Proc. R. Soc. Lond. B* **264**, 1287–1291 (1997).
10. Fedtke, I., Gotz, F. & Peschel, A. *Int. J. Med. Microbiol.* **294**, 189–194 (2004).

SEISMOLOGY

The start of something big?

Rachel Abercrombie

Can we predict the final size of an earthquake from observations of its first few seconds? An extensive study of earthquakes around the Pacific Rim seems to indicate that we can — but uncertainties remain.

How does a seismic fault, initially essentially immobile, start to slip at speeds of metres per second as an earthquake rupture front runs along it at speeds of up to 3 kilometres per second? Does the eventual size of an earthquake depend on the nature of this process? Or do all earthquakes begin in the same way, with the extent of rupture determined by conditions along the fault? Such fundamental questions get seismologists talking, because knowing how earthquakes begin is an essential part of understanding and modelling the dynamics of earthquake rupture, and may allow an earthquake's course to be predicted. Research until now has been inconclusive, but results described by Olson and Allen (page 212 of this issue)¹ imply that the final magnitude of an

earthquake depends at least partially on what happens in its first few seconds. This timescale is equivalent to less than a tenth of the duration of the larger earthquakes in their study.

Research into the onset of earthquakes large and small has found that they often begin with small-amplitude shaking². The interpretation of these initial 'sub-events' remains controversial. One model has it that a small, isolated sub-event triggers a larger fault patch, which itself triggers further fault patches, and so on as long as sufficient energy is available. In this 'cascade' model, the beginning of a large earthquake is no different from the beginning of a small earthquake: therefore, predicting the final magnitude from the first few seconds is impossible. An alternative model is that the small



Figure 1 | Finding fault. A view of the Lavic Lake seismic fault in California. The Hector Mine earthquake, one of those considered by Olson and Allen¹ in their study of the initial waves of Pacific Rim earthquakes, occurred along this fault line on 16 October 1999.

K. M. JOHNS/SPL

beginning is the last phase of some longer, slower, sub-seismic 'nucleation' process. (Such a process has admittedly never been reliably observed³.) In this case, the final magnitude would be related to the nature of the nucleation process, and seismograms of large earthquakes would look different from those of smaller earthquakes right from the start.

Early warning systems currently in operation in Japan, Taiwan and Mexico use observations of the earliest-arriving primary (P) waves to provide a few seconds' warning of subsequent large ground motion — secondary (S) and surface waves — produced by the same earthquake. In an earlier study⁴, Allen and Kanamori investigated the first few seconds of earthquake seismograms in southern California. They found that the predominant period (a measure of the frequency) for the first 4 seconds of the P waves provides a good estimate of the size of earthquakes with a magnitude M of less than 6. The duration of such earthquakes, defined as the time during which the fault actually moves, is usually less than 4 seconds (the waves generated by an earthquake last for much longer than the earthquake itself). Intriguingly, however, Allen and Kanamori's method⁴ also predicted the approximate magnitude of three earthquakes of M greater than 6, and so an earthquake duration of more than 4 seconds. In other words, the final size of the earthquake could be predicted before the fault stopped moving.

Olson and Allen¹ set out to determine whether the final magnitude of the earthquake really does depend on the predominant period of the onset. They investigated the first few seconds of 71 earthquakes from California, Alaska, Japan and Taiwan, each recorded at multiple stations within 100 kilometres of the epicentres. Twenty-four of the earthquakes had a magnitude larger than 6, with durations of up to 70 seconds. Estimating the predominant period of the radiated seismic energy for each earthquake, the authors find that this value increases with magnitude for earthquakes of M between 3 and 8. This finding applies even to larger earthquakes in which the measurement is made after as little as a tenth of the earthquake's total duration — suggesting that the final magnitude of an earthquake is indeed determined a very short time after onset.

Previous studies of earthquake onsets have been limited by the lack of seismometers located close to the epicentre, and by the fact that standard techniques cannot analyse the frequency content of such short pieces of the seismograms. The method⁵ used by Olson and Allen¹, and by Allen and Kanamori before them⁴, is simple but effective. They calculate the predominant period from the ratio of the ground displacement to the rate of change of that displacement (the velocity of the movement) point by point. This measurement can be made as a seismogram is recorded, and at seismometers up to

100 kilometres from an earthquake's epicentre.

As Olson and Allen note, there is considerable scatter in their results; this leads to large uncertainties, especially in measurements at individual seismometers. An individual measurement of a predominant period of 1 second, for example, is consistent with an earthquake of any magnitude between 3 and 7.5. Most measurements of earthquake parameters vary significantly between seismometers, but even after averaging over many stations, any measurement of the mean predominant period produces an uncertainty of at least one magnitude unit. The predominant period of the 1999 earthquake in Hector Mine, California ($M = 7.1$; Fig. 1), for instance, is the same as that of an earthquake of M less than 5.

The relationship between the first 4 seconds of an earthquake and its final magnitude implies either that there is an initial, sub-seismic nucleation phase that is proportional to the size of the earthquake, or that any triggering cascade of sub-events lasts less than

4 seconds (the approximate duration of an earthquake of $M = 6$). But these observations of earthquake onsets are purely empirical, and we are far from understanding how onset, propagation and state of stress of the surrounding fault interact to determine the final size of a seismic event. Olson and Allen's study advances that understanding, and thus our ability to predict an earthquake's size before it reaches its peak. It also raises intriguing questions worthy of further study. ■

Rachel Abercrombie is in the Department of Earth and Planetary Sciences, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts 02215, USA.
e-mail: rea@bu.edu

1. Olson, E. L. & Allen, R. M. *Nature* **438**, 212–215 (2005).
2. Ellsworth, W. L. & Beroza, G. C. *Science* **268**, 851–855 (1995).
3. Bakun, W. H. *et al.* *Nature* **437**, 969–974 (2005).
4. Allen, R. M. & Kanamori, H. *Science* **300**, 786–789 (2003).
5. Nakamura, Y. in *Proc. 13th World Conf. Earthquake Eng.* Pap. No. 908 (2004).

CIRCADIAN RHYTHMS

Clock coordination

Michael N. Nitabach

Many animals concentrate their activity around dawn and dusk. This timing is regulated by distinct 'morning' and 'evening' oscillators in the central nervous system. But how are these two neuronal clocks coordinated?

A diverse range of organisms, from algae to human beings, have internal clocks that are synchronized to Earth's 24-hour rotation. When deprived of environmental cues, for example by being kept in constant darkness, these clocks 'free-run'; that is, they independently maintain rhythms of roughly 24 hours that gradually drift out of synchrony with Earth's rotation (hence their designation as circadian clocks). In animals, the clocks that control the timing of rest and activity are embodied in groups of neurons in the central nervous system.

One of the big questions in this field is how the oscillations of the multiple cellular clocks within an organism are coordinated. For example, the fruitfly *Drosophila melanogaster* possesses two anatomically distinct groups of clock neurons that can independently control its morning and evening peaks of activity^{1,2}. Under normal conditions, however, the morning (M) and evening (E) oscillators run coordinately. On page 238 of this issue, Stoleru *et al.*³ show that the M oscillator sends a daily resetting signal to the E oscillator to keep the two in synchrony.

Almost 30 years ago, it was theorized that nocturnal rodent circadian rhythms are controlled by independent, but coupled, M and E oscillators⁴. Later measurements of neuron

firing in the rodent brain nucleus that contains the clock cells revealed two distinct subpopulations of neurons whose firing rhythms were out of phase, suggesting that these could be the M and E oscillators⁵. As with some nocturnal rodents, diurnal fruitflies show peaks of activity centred around the transitions from night to day and from day to night.

In previous work, Stoleru and colleagues¹ generated fruitflies that lacked either the lateral-ventral anatomical group of clock neurons or the dorsal group. Fruitflies lacking the lateral-ventral subgroup lost the morning peak of activity, but retained the evening peak. In fruitflies lacking the dorsal subgroup the reverse was true. Along with the results of Grima *et al.*², published simultaneously, these findings localized the M and E oscillators to the lateral-ventral and dorsal clock neurons, respectively.

Those experiments were performed in 12-hour:12-hour light:dark (LD) conditions, in which both oscillators can be independently synchronized to the environment and thereby maintain a constant phase relationship. When fruitflies are synchronized to LD cycles and then released into constant darkness (DD), the morning and evening peaks still occur — although they free-run and so gradually drift out of phase with the rotation of Earth. But

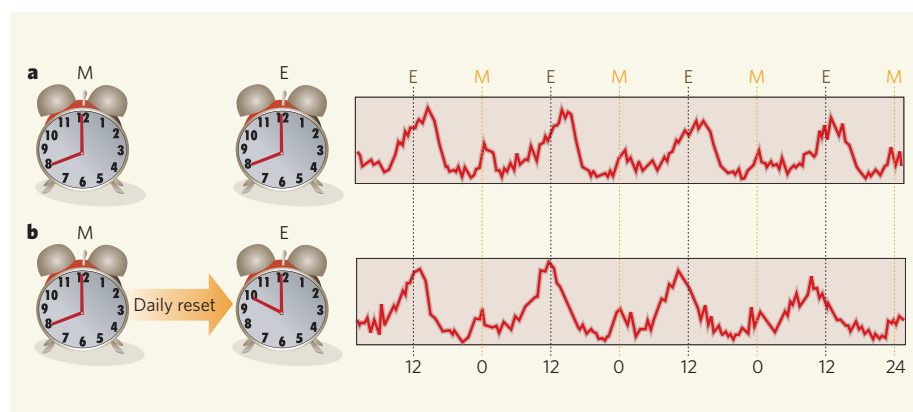


Figure 1 | Synchronization of fruitfly circadian oscillators. The oscillators in the fruitfly (*Drosophila melanogaster*) brain that control morning (M) and evening (E) peaks of locomotor activity reside in distinct neuronal cell populations. **a**, Under normal circumstances, the M and E oscillators run with the same intrinsic period. **b**, When the E oscillator is genetically modified to run with an intrinsic period shorter than that of the M oscillator, the evening peak of activity occurs earlier than usual. But the peak-to-peak period of the evening peak of activity is the same as that of the morning peak. This suggests that a signal is sent to the E oscillator once each day by the M oscillator to synchronize the phase of E to that of M, whereas in between these resets the E oscillator runs with its intrinsic period.

even in free-running DD conditions, the morning and evening peaks maintain a constant phase relationship with each other. So there must be some mechanism for keeping that relationship even in the absence of any environmental cues.

Stoleru *et al.* tackled this mechanism in their new paper³. The first question they addressed is whether there is a master–slave relationship between the M and E oscillators, or whether these oscillators mutually influence each other to collaboratively arrive at a constant phase relationship. To do so, the authors created genetically modified fruitflies in which the M and E oscillators run with intrinsic periods that differ by three to four hours. They found that, under free-running DD conditions, the period of behavioural locomotor activity is always determined by the intrinsic period of the M oscillator, suggesting that there is a master–slave relationship and that M is the master.

Moving to the cellular level, Stoleru *et al.* determined the period of the clock in M and E oscillator cells using a molecular marker. As would be expected if M is the master and E the slave, the period of clock oscillation in the E cells was determined by that of the M cells, and not by the intrinsic period of the E cells themselves.

So how does M exercise control? One possibility is that it exerts a continuous drive on E such that cellular oscillations in E track those of M on an hour-by-hour basis. Alternatively, it could be that E free-runs with its intrinsic period but is reset once a day by M so that its overall period is the same as M's.

To distinguish between these possibilities, Stoleru *et al.*³ compared the phase of the evening peak of locomotor activity in fruitflies in which E had a genetically induced short intrinsic period and a normal M, with that in fruitflies with E and M both of normal period. The result was that fruitflies in which E runs fast exhibit their evening peak of activity about

two hours earlier than normal fruitflies (Fig. 1). However, the overall peak-to-peak period of E is the same as that of M, suggesting that the M oscillator resets the E oscillator once a day. In between these resets E free-runs with its own faster intrinsic period.

Research does not end here, of course. Most obviously there is the issue of the resetting signal that M sends to E, and here one can speculate about the involvement of a neuropeptide called pigment-dispersing factor (PDF). M cells express PDF, and, in the absence of PDF signalling, the evening peak in LD conditions occurs about two hours earlier than in normal fruitflies⁶. Moreover, the PDF receptor has recently been identified, and it is indeed expressed in at least some of the cells that comprise the E oscillator^{7–9}. Closer to home, investigators will be naturally curious as to whether the M and E oscillators in the mammalian brain communicate in a similar fashion to those in the fruitfly. It is to be expected that future work will resolve these and other fascinating questions raised by Stoleru and colleagues' findings.

Michael N. Nitabach is in the Department of Cellular and Molecular Physiology, Interdepartmental Neuroscience Program, Yale School of Medicine, New Haven, Connecticut 06520, USA.
e-mail: michael.nitabach@yale.edu

1. Stoleru, D., Peng, Y., Agosto, J. & Rosbash, M. *Nature* **431**, 862–868 (2004).
2. Grima, B., Chelot, E., Xia, R. & Rouyer, F. *Nature* **431**, 869–873 (2004).
3. Stoleru, D., Peng, Y., Nawathean, P. & Rosbash, M. *Nature* **438**, 238–242 (2005).
4. Pittendrigh, C. S. & Daan, S. *J. Comp. Physiol.* **106**, 333–355 (1976).
5. Jagota, A., de la Iglesia, H. O. & Schwartz, W. J. *Nature Neurosci.* **3**, 372–376 (2000).
6. Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C. & Taghert, P. H. *Cell* **99**, 791–802 (1999).
7. Mertens, I. *et al. Neuron* **48**, 213–219 (2005).
8. Lear, B. C. *et al. Neuron* **48**, 221–227 (2005).
9. Hyun, S. *et al. Neuron* **48**, 267–278 (2005).



50 YEARS AGO

"New Records in Human Power"

— When testing the Swedish national ski-team for physical condition...values for oxygen intake as high as 6.1 l./min. were estimated. As these figures far surpassed the earlier maximum, it was of great interest to obtain direct measurements of those subjects' aerobic capacity... values are included from an investigation of capacity for hard muscular work of eighty-six healthy, well-trained students of physical education... The most striking difference between those students, on one hand, and the middle-distance runners and skiers, on the other, is found when comparing the capacity for oxygen uptake. Values for vital capacity, maximum heart-rate, pulmonary ventilation and concentration of blood lactic acid were of the same order. It is probable that a high aerobic capacity is an essential characteristic of people with a high standard of physical fitness with regard to endurance.

From *Nature* 12 November 1955.

100 YEARS AGO

Science scholars selected from the whole of Great Britain for their ability and promise, maintaining themselves on 17s. 9d. per week, were this year saved from much privation by secret gifts of small bursaries — see subjoined audited account. I have no right to ask for help from the generous men who helped me last year, but I have all the sturdiness of a chartered beggar — I ask in a good cause.

It was originally intended that these bursaries should be given only to such National Scholars as required assistance, but some of the subscribers have given me power to assist other students of the college. Also one of the two City Companies has given me power to grant an occasional bursary of more than ten pounds. It is understood that every student is morally bound to repay this money to the fund at some future time.

John Perry

From *Nature* 9 November 1905.

50 & 100 YEARS AGO

BRIEF COMMUNICATIONS

Gravitational tractor for towing asteroids

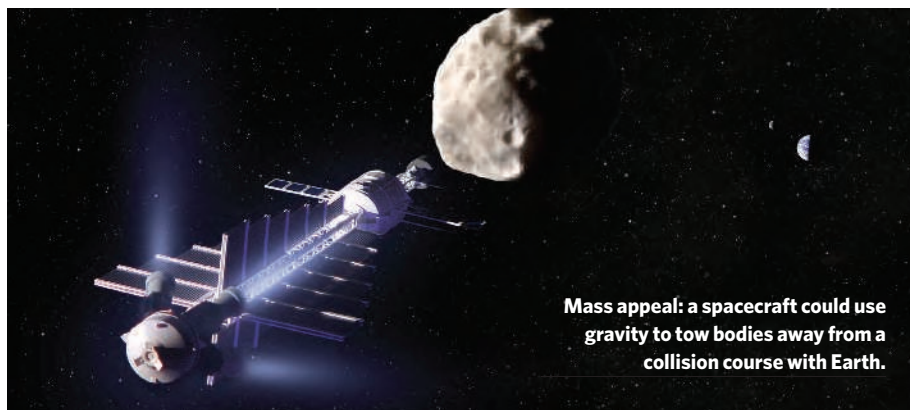
A spacecraft could deflect an Earth-bound asteroid without having to dock to its surface first.

We present a design concept for a spacecraft that can controllably alter the trajectory of an Earth-threatening asteroid by using gravity as a towline. The spacecraft hovers near the asteroid, with its thrusters angled outwards so that the exhaust does not impinge on the surface. This proposed deflection method is insensitive to the structure, surface properties and rotation state of the asteroid.

The collision of a small asteroid of about 200 m with the Earth could cause widespread damage and loss of life¹. One way to deflect an approaching asteroid is to dock a spacecraft to the surface and push on it directly². The total impulse needed for rendezvous and deflection is too large for chemical rockets, but would be achievable by spacecraft such as the 20-tonne nuclear-electric propelled vehicles that were proposed as part of NASA's Prometheus programme².

Regardless of the propulsion scheme, a docked asteroid tug needs an attachment mechanism because the surface gravity is too weak to hold it in place. Asteroids are likely to be rough and unconsolidated, making stable attachment difficult. Furthermore, most asteroids rotate, so an engine anchored to the surface thrusts in a constantly changing direction. Stopping the asteroid's rotation, reorienting its spin axis³, or firing the engine only when it rotates through a certain direction, adds complexity and wastes time and propellant.

Our suggested alternative is to have the spacecraft simply hover above the surface of the asteroid. The spacecraft tows it without physical attachment by using gravity as a towline. The thrusters must be canted outboard to keep them from blasting the surface (which



Mass appeal: a spacecraft could use gravity to tow bodies away from a collision course with Earth.

D. DURDA, FIAAA/BGIZ FOUNDATION

would reduce the net towing force and stir up unwanted dust and ions).

This scheme is insensitive to the poorly understood surface properties, internal structures and rotation states of asteroids. A spacecraft needs only to keep its position in the direction of towing while the target asteroid rotates beneath it. The engines must be actively throttled to control the vertical position as the equilibrium hover point is unstable. The horizontal position is controlled by differential throttling of engines on opposite sides of the spacecraft. The spacecraft can be made stable in attitude by designing it like a pendulum, with the heaviest components hanging closest to the asteroid and the engines farther away.

The thrust required to balance the gravitational attraction is given by

$$T \cos[\sin^{-1}(r/d) + \phi] = GMm/d^2 \\ = 1.7 \left(\frac{r}{d} \right)^3 \left(\frac{\rho}{2 \times 10^3} \right) \left(\frac{m}{20 \times 10^3} \right) \left(\frac{d}{150} \right)$$

where G is the gravitational constant; see Fig. 1 for definition of other variables. Thus a 20-tonne spacecraft with $\phi = 20^\circ$ hovering one half-radius above the surface ($d/r = 1.5$) can tow an asteroid of 200 m diameter and density $\rho = 2 \times 10^3 \text{ kg m}^{-3}$, provided it can maintain a total thrust T of just over 1 newton.

The velocity change imparted to the asteroid per second of hovering (Δv) is given by

$$\Delta v = \frac{Gm}{d^2} = 5.9 \times 10^{-11} \left(\frac{m}{20 \times 10^3} \right) \left(\frac{150}{d} \right)^2$$

So the velocity change imparted to the asteroid in our example in a single year of hovering is $1.9 \times 10^{-3} \text{ m s}^{-1}$. Because Δv is largely independent of the asteroid's detailed structure and composition, the effect on the asteroid's orbit

is predictable and controllable, as would be required for a practical deflection scheme.

The mean change in velocity required to deflect an asteroid from an Earth impact trajectory is about $3.5 \times 10^{-2} t \text{ m s}^{-1}$, where t is the lead time in years⁴. So a 20-tonne gravitational tractor hovering for one year can deflect a typical asteroid of about 200 m diameter given a lead time of roughly 20 years.

The thrust and total fuel requirements of our mission example would be well within the capability of proposed 100-kilowatt nuclear-electric propulsion systems², using about 4 tonnes of fuel to accomplish the typical 15 km s^{-1} rendezvous and about 400 kg for the actual deflection. For a given spacecraft mass, the fuel required for the deflection scales linearly with the asteroid mass.

Deflecting a larger asteroid would require a heavier spacecraft, more time spent hovering, or more lead time. However, in the special case in which an asteroid has a close Earth approach, followed by a later return and impact, the change in velocity needed to prevent the impact can be many orders of magnitude smaller if applied before the close approach⁵. For example, the asteroid 99942 Apophis (2004 MN4), a 320-m asteroid that will swing by the Earth at a distance of about 30,000 km in 2029, has a small probability (10^{-4}) of returning to strike the Earth in 2035 or 2036 (ref. 6). If it is indeed on a return impact trajectory, a deflection of only about 10^{-6} m s^{-1} a few years before the close approach in 2029 would prevent a later impact (A. Carusi, personal communication). In this case, a 1-tonne gravitational tractor with conventional chemical thrusters could accomplish this deflection mission as only 0.1 newtons of thrust would be required for a duration of about a month. Should such a deflection mission

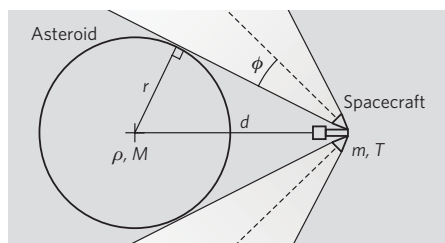


Figure 1 | Towing geometry of a gravitational tractor. The asteroid (assumed to be spherical) has radius r , density ρ and mass M . The spacecraft has mass m , total thrust T and an exhaust-plume half-width ϕ . It hovers at distance d from the asteroid's centre, where its net thrust balances its weight. The thrusters are tilted outwards to prevent exhaust impinging on the asteroid surface.

prove necessary, a gravitational tractor offers a viable method of controllably steering asteroid 99942 Apophis away from an Earth impact.

Edward T. Lu, Stanley G. Love

NASA Johnson Space Center, Mail Code CB,
Houston, Texas 77058, USA
e-mail: edward.t.lu@nasa.gov

1. Chapman, C. R. *Earth Planet. Sci. Lett.* **222**, 1–15 (2004).
2. Schweickart, R. L., Lu, E. T., Hut, P. & Chapman, C. R. *Sci. Am.* **289**, 54–61 (2003).

3. Scheeres, D. J. & Schweickart, R. L. *The Mechanics of Moving Asteroids* (paper 2004-1446, Am. Inst. Aeronaut. Astronaut., 2004).
4. Chesley, S. R. & Spahr, T. B. in *Mitigation of Hazardous Comets and Asteroids* (eds Belton, M. J. S. et al.) 22–37 (Cambridge Univ. Press, Cambridge, 2004).
5. Carusi, A., Valsecchi, G. B., D'Abramo, G. & Boatini, A. *Icarus* **159**, 417–422 (2002).
6. JPL Sentry Impact Risk Page <http://neo.jpl.nasa.gov/risk>

Competing financial interests: declared none.
doi:10.1038/438177a

GREEN CHEMISTRY

Biodiesel made with sugar catalyst

The production of diesel from vegetable oil calls for an efficient solid catalyst to make the process fully ecologically friendly. Here we describe the preparation of such a catalyst from common, inexpensive sugars. This high-performance catalyst, which consists of stable sulphonated amorphous carbon, is recyclable and its activity markedly exceeds that of other solid acid catalysts tested for 'biodiesel' production.

The esterification of higher fatty acids by liquid acid catalysts such as sulphuric acid (H_2SO_4) is a process commonly used for biodiesel production, but it involves high consumption of energy and the separation of the catalysts from the homogeneous reaction mixtures is costly and chemically wasteful. Recyclable solid acids, such as Nafion^{1–4}, make better catalysts, although they are also expensive and their activity is less than that of liquid acids¹. Sulphonated naphthalene carbonized at 200–250 °C is a solid acid catalyst that has been used successfully for ethyl acetate formation⁵; however, it is a soft material and its aromatic molecules are leached out during liquid-phase reactions above 100 °C or when higher fatty acids are used as surfactants, so its catalytic activity is rapidly lost.

We have devised a strategy to overcome these problems by sulphonating incompletely carbonized natural organic material to prepare a more robust solid catalyst. Incomplete carbonization of natural products such as sugar, starch or cellulose results in a rigid carbon material that is composed of small polycyclic

aromatic carbon sheets in a three-dimensional sp^3 -bonded structure. Sulphonation of this material would be expected to generate a stable solid with a high density of active sites, enabling a high-performance catalyst to be prepared cheaply from naturally occurring molecules.

The scheme we use to sulphonate incompletely carbonized saccharides is shown in Fig. 1. First, D-glucose and sucrose are incompletely carbonized at low temperature to induce pyrolysis and the formation of small polycyclic aromatic carbon rings; sulphonite groups ($-\text{SO}_3\text{H}$) are then introduced by sulphuric acid (see supplementary information). Structural analysis^{6–8} indicates that the prepared samples consist of sheets of amorphous carbon bearing hydroxyl and carboxyl ($-\text{OH}$ and $-\text{COOH}$) groups, as well as high densities of $-\text{SO}_3\text{H}$ groups.

This black powder is insoluble in water, methanol, benzene, hexane, *N,N*-dimethylformamide and oleic acid, even at boiling temperatures. It can be moulded into hard pellets or thin flexible films by heating with 0.5–5.0% by weight of binding polymer; the two forms have comparable stability and catalytic performance. The thin films act as electrically insulating proton conductors whose properties (0.09 siemens cm^{-1} at 50 °C and 100% humidity) are comparable to that of Nafion (0.1 siemens cm^{-1} at 80 °C).

High-grade biodiesel is produced by esterification of the vegetable-oil constituents oleic acid and stearic acid. The activity of our solid sulphonated carbon catalyst in this reaction is

more than half that of a liquid sulphuric acid catalyst and much higher than can be achieved by conventional solid acid catalysts (see supplementary information). There was no loss of activity or leaching of $-\text{SO}_3\text{H}$ during the process, even for samples subjected to repeated reactions at 80–180 °C after having been recovered by simple decantation. The activity is double that of a carbonized sulphonated naphthalene catalyst tested previously⁵, which decreased rapidly on recycling at 80 °C.

Carbon catalysts identical to those described here have also been successfully produced from carbonized starch and cellulose (results not shown). Saccharide molecules may therefore be generally suitable for preparing these catalysts, which can be used as a replacement for liquid sulphuric acid in esterification reactions. In addition to biodiesel production, such environmentally benign alternative catalysts should find application in a wide range of other acid-catalysed reactions.

**Masakazu Toda*, Atsushi Takagaki*,
Mai Okamura*, Junko N. Kondo*,
Shigenobu Hayashi†, Kazunari Domen‡,
Michikazu Hara***

*Chemical Resources Laboratory, Tokyo Institute of Technology, Yokohama 226-8503, Japan
e-mail: mhara@res.titech.ac.jp

†Research Institute of Instrumentation Frontier, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8565, Japan

‡Department of Chemical System Engineering, School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan

1. Okuhara, T. *Chem. Rev.* **102**, 3641–3666 (2002).
2. Clark, J. H. *Acc. Chem. Res.* **35**, 791–797 (2002).
3. Misono, M. *C. R. Acad. Sci. II: Chimie* **3**, 471–475 (2000).
4. Smith, K., El-Hiti, G. A., Jayne, A. J. & Butters, M. *Org. Biomol. Chem.* **1**, 1560–1564 (2003).
5. Hara, M. et al. *Angew. Chem. Int. Edn* **43**, 2955–2958 (2004).
6. Tsubouchi, N., Xu, K. & Ohtsuka, Y. *Energy Fuels* **17**, 1119–1125 (2003).
7. Silverstein, R. M., Bassler, G. C. & Morrill, T. C. in *Spectrometric Identification of Organic Compounds* 5th edn 218–232 (Wiley, Indianapolis, 1991).
8. Zhang, X. & Solomon, D. H. *Chem. Mater.* **11**, 384–391 (1999).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/438178a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

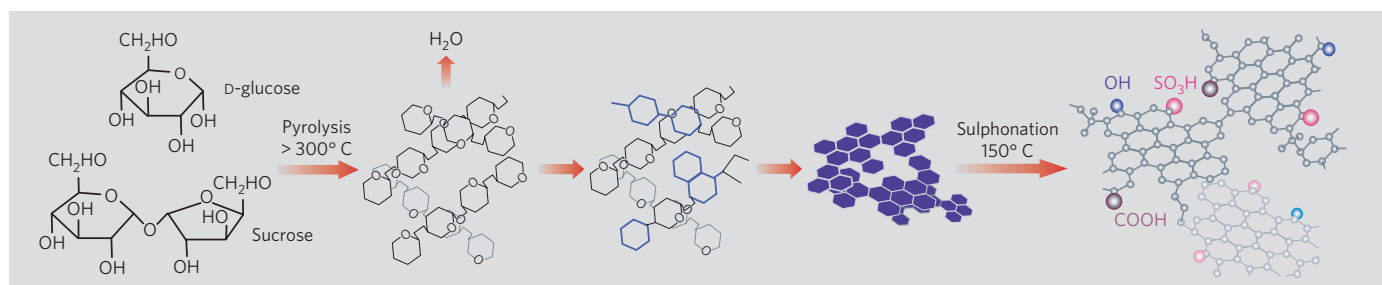


Figure 1 | Preparation from sucrose and D-glucose of a solid catalyst suitable for biological diesel production. Pyrolysis of the sugars causes their incomplete carbonization (middle; outlined in blue) and formation into polycyclic aromatic carbon sheets; sulphuric acid (concentrated or fuming) is used to sulphonate the aromatic rings to produce the catalyst. For details of methods, see supplementary information.

NEUROSCIENCE

Rewiring the adult brain

Arising from: S. M. Smirnakis *et al.* *Nature* **435**, 300–307 (2005)

Any analysis of plastic reorganization at a neuronal locus needs a veridical measure of changes in the functional output — that is, spiking responses of the neurons in question. In a study of the effect of retinal lesions on adult primary visual cortex (V1), Smirnakis *et al.*¹ propose that there is no cortical reorganization. Their results are based, however, on BOLD (blood-oxygen-level-dependent) fMRI (functional magnetic resonance imaging), which provides an unreliable gauge of spiking activity. We therefore question their criterion for lack of plasticity, particularly in the light of the large body of earlier work that demonstrates cortical plasticity.

Plasticity in adult V1 has been demonstrated by multiple, independent lines of evidence from more than twenty studies in three species (see refs 2–6, for example). Physiologically, the evidence derives from measurements of lesion-induced shifts in the locations of V1 neuronal receptive fields. By plotting receptive fields before and at various points after making retinal lesions, it was shown that the affected cortex — with receptive fields originally inside the lesion — develops new, shifted receptive field positions after recovery. These shifts are cortically mediated because the lateral geniculate nucleus, the source of thalamic input to V1, shows limited reorganization⁷. All these measurements were made using suprathreshold spiking neuronal responses: this is an important point as the neuronal output from V1 to subsequent cortical stages is carried entirely by spikes. Any measure of V1 reorganization — and consequent functional remapping of visual information — therefore needs to assess the effect on spiking activity. These physiological results are buttressed by anatomical findings showing a selective increase in the density of axon collaterals in reorganized cortex⁸, and the sequential expression of biochemical markers^{9,10}.

By contrast, the primary evidence for lack of plasticity offered by Smirnakis *et al.* is the observation that the V1 ‘silent zone’, mapped with BOLD fMRI immediately following a retinal lesion, did not change over time. There are plausible reasons why fMRI maps may fail to change, despite re-emergent neuronal activity. The reorganization of cortex is believed to be mediated by long-range horizontal connections within V1 (refs 8, 11, 12). In normal V1, these connections mediate subthreshold modulation. Following retinal lesions, horizontal connections stretching from ‘normal’ cortex into the lesion projection zone (LPZ) are believed to strengthen their synapses — but not to change anatomical extent. They there-

fore induce re-emergent spiking activity, but only in neurons lying within their target zone in the silenced cortex.

Such re-emergent activity could involve reduction of inhibition as much as an increase in excitation. As the BOLD signal probably reflects synaptic input into a region rather than spiking output, the ‘silent zone’ observed by Smirnakis *et al.* immediately following a lesion may mark not the edge of the real LPZ but the inner edge of subthreshold activation spreading into the LPZ through horizontal connections. In subsequent measurements, the BOLD signal would continue to show the unchanging position of this inner boundary while being blind to synaptic reorganization, which would lead to re-emergent spiking activity over the extent of the horizontal connections. The single set of electrode recordings by Smirnakis *et al.* after months of recovery might simply show the extent of largely completed recovery and, not surprisingly, produce a border in register with the edge of the BOLD signal.

Furthermore, BOLD gives a local measure of the total cortical activity, a significant component of which comes from thalamocortical inputs, the contribution of which is probably further accentuated by the disproportionately high vascularization of layer 4, the cortical input layer. However, the neurons showing recovery may reside primarily in the superficial layers¹², which receive the long-range horizontal connections, as opposed to layer 4.

Owing to these uncertainties about the validity of BOLD fMRI as a yardstick of functional reorganization in V1, we believe that Smirnakis *et al.* do not present a convincing contradiction to the body of earlier evidence indicating substantial receptive field plasticity in adult animals following retinal lesion. The recovered activity demonstrated in the earlier studies has a likely corollary in the recovery of visual perception: human subjects suffering

from macular degeneration, or with artificially induced retinal lesions, show improved perceptual fill-in over time after the lesions^{13–15}.

Michael B. Calford*, Yuzo M. Chino†, Aniruddha Das‡, Ulf T. Eysel§, Charles D. Gilbert||, Stephen J. Heinen¶, Jon H. Kaas#, Shimon Ullman☆

*School of Biomedical Sciences, The University of Newcastle, New South Wales 2308, Australia

†College of Optometry, University of Houston, Houston, Texas 77004, USA

‡Columbia University Medical School, New York, New York 10032, USA

§Department of Neurophysiology, Faculty of Medicine, Ruhr-University Bochum, 44780 Bochum, Germany

||The Rockefeller University, New York, New York 10021, USA

¶The Smith-Kettlewell Eye Research Institute, San Francisco, California 94115, USA

#Vanderbilt University, Nashville, Tennessee 37240, USA

☆Weizmann Institute, Rehovot 76100, Israel

1. Smirnakis, S. M. *et al.* *Nature* **435**, 300–307 (2005).
2. Heinen, S. J. & Skavenski, A. A. *Exp. Brain Res.* **83**, 670–674 (1991).
3. Calford, M. B. *et al.* *J. Physiol. Lond.* **524**, 587–602 (2000).
4. Gilbert, C. D., Hirsch, J. A. & Wiesel, T. N. *Cold Spring Harbor Symp. Quant. Biol.* **55**, 663–677 (1990).
5. Kaas, J. H. *et al.* *Science* **248**, 229–231 (1990).
6. Chino, Y. M., Smith, E. L. III, Kaas, J. H., Sasaki, Y. & Cheng, H. J. *Neurosci.* **15**, 2417–2433 (1995).
7. Eysel, U. T. *Nature* **299**, 442–444 (1982).
8. Darian-Smith, C. & Gilbert, C. D. *Nature* **368**, 737–740 (1994).
9. Obata, S., Obata, J., Das, A. & Gilbert, C. D. *Cereb. Cortex* **9**, 238–248 (1999).
10. Arckens, L. *et al.* *Eur. J. Neurosci.* **12**, 4222–4232 (2000).
11. Das, A. & Gilbert, C. D. *Nature* **375**, 780–784 (1995).
12. Calford, M. B., Wright, L. L., Metha, A. B. & Tagliaventi, V. *J. Neurosci.* **23**, 6434–6442 (2003).
13. Craik, K. J. W. in *The Nature of Psychology* (ed. Sherwood, S. L.) 98–103 (Cambridge Univ. Press, Cambridge, 1966).
14. Gerrits, H. J. & Timmerman, G. J. *Vision Res.* **9**, 439–442 (1969).
15. Zur, D. & Ullman, S. *Vision Res.* **43**, 971–982 (2003).

doi:10.1038/nature04359

NEUROSCIENCE

Smirnakis *et al.* reply

Replying to: M. B. Calford *et al.* *Nature* **438**, doi:10.1038/04359 (2005)

We disagree with Calford *et al.*¹ that there is a consensus on adult plasticity in primate V1 cortex: for example, macaque area V1 cytochrome oxidase levels remained depressed for several months after binocular retinal lesions²; no reorganization in macaque V1 after monocular retinal lesions was found³; and no area

V1 reorganization in a patient with macular degeneration was detected⁴.

Calford *et al.*¹ agree that subthreshold activity shows no long-term reorganization. They propose that plasticity comprises only an increase in the likelihood of transforming subthreshold signals into action

potentials. Their argument hinges on the claim that even our first BOLD (blood-oxygen-level-dependent) maps, obtained immediately after the lesion, reflect the full extent of reorganization (assumed to be equal in extent to sub-threshold synaptic activity). Our results do not support this view. Rather, the BOLD-defined lesion projection zone (LPZ) border matches its anticipated location based on the retinotopic projection of the retinal lesion. Specifically, the BOLD-defined LPZ borders in Figs 1,2 of ref. 5 are within 1.5 mm of their expected retinotopic projection. This is commensurate with the expected vascular spread and considerably smaller than most estimates of reorganization (about 2–5 mm).

Calford *et al.*¹ defend a weak definition of plasticity. The prevalent view of long-term reorganization involves creation of new connections and increased efficacy of existing synapses⁶, which is expected⁷ to increase sub-threshold activity. However, we observed no change in the slope or position of the BOLD profile at the LPZ border, suggesting that changes in subthreshold activity are weak or occur over a small scale (1 mm). Even if neuronal firing increases by down-regulation of inhibition, as proposed by Calford *et al.*, the BOLD signal in the neocortex should increase⁸. We note that functional magnetic resonance imaging is sensitive for detecting reorganization in human cortex^{9,10}, and that BOLD is strongest in the upper 1 mm of cortex¹¹, where long-range horizontal connections reside.

Calford *et al.* suggest that receptive field shifts at the LPZ border unequivocally demonstrate long-term plasticity. This is not neces-

sarily so. Retinal recovery from thermal injury at the lesion border could produce similar shifts². Moreover, the time course of this phenomenon is unclear, with some studies^{12,13} reporting large (up to 4.5 mm), almost immediate activity shifts (discussed in ref. 5). Neither does perceptual 'filling-in' necessarily imply V1 reorganization. 'Filling-in' has been proposed to occur in higher areas¹⁴, and macaque perceptual 'filling-in' has been demonstrated without area V1 reorganization³. Although anatomical and immunohistochemical changes occur inside the LPZ after deafferentiation^{6,15}, their functional significance is unclear.

Our multi-unit recordings from superficial and deep cortical layers confirm the BOLD measurements, showing that steady-state responses do not recover substantially inside the LPZ, and ruling out the possibility that BOLD missed reorganization by reflecting primarily activity in the thalamo-cortical afferents⁵.

By monitoring aggregate neural activity, we may have missed reorganization in select neuronal subpopulations. As discussed⁵, we cannot exclude short-term (minutes to hours) reorganization, or a limited form of plasticity expressed as adaptive gain adjustments affecting selectively transient extra-classical responses inside the LPZ while leaving steady-state responses largely unchanged. No method could convincingly show a complete lack of changes after deafferentiation, but our measurements define limits on the strength and spatial extent of any proposed reorganization.

Stelios M. Smirnakis*†, Michael C. Schmid*,

Alyssa A. Brewer‡, Andreas S. Tolias*, Almut Schüz*, Mark Augath*, Werner Inhoffen§, Brian A. Wandell‡, Nikos K. Logothetis*

*Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

e-mail: smsmirnakis@partners.org

†Department of Neurology, Massachusetts General Hospital and Brigham and Women's Hospital, Harvard University, Boston, Massachusetts 02114, USA

‡Neuroscience Program and Department of Psychology, Stanford University, Stanford, California 94305, USA

§Department of Ophthalmology I, University of Tübingen, 72076 Tübingen, Germany

1. Calford, M. B. *et al.* *Nature* doi:10.1038/nature04359 (2005).
2. Horton, J. C. & Hocking, D. R. *J. Neurosci.* **18**, 5433–5455 (1998).
3. Murakami, I., Komatsu, H. & Kinoshita, M. *Vis. Neurosci.* **14**, 89–101 (1997).
4. Sunness, J. S., Liu, T. & Yantis, S. *Ophthalmology* **111**, 1595–1598 (2004).
5. Smirnakis, S. M. *et al.* *Nature* **435**, 300–307 (2005).
6. Darian-Smith, C. & Gilbert, C. D. *Nature* **368**, 737–740 (1994).
7. Das, A. & Gilbert, C. D. *Nature* **375**, 780–784 (1995).
8. Chen, Z., Silva, A. C., Yang, J. & Shen, J. *J. Neurosci. Res.* **79**, 383–391 (2005).
9. Baseler, H. A., Morland, A. B. & Wandell, B. A. *J. Neurosci.* **19**, 2619–2627 (1999).
10. Baker, C. I., Peli, E., Knouf, N. & Kanwisher, N. G. *J. Neurosci.* **25**, 614–618 (2005).
11. Kennerley, A. J. *et al.* *Magn. Reson. Med.* **54**, 354–365 (2005).
12. Pettet, M. W. & Gilbert, C. D. *Proc. Natl Acad. Sci. USA* **89**, 8366–8370 (1992).
13. Schmid, L. M., Rosa, M. G. & Calford, M. B. *Neuroreport* **6**, 1349–1353 (1995).
14. De Weerd, P., Gattass, R., Desimone, R. & Ungerleider, L. G. *Nature* **377**, 731–734 (1995).
15. Hendry, S. H. & Jones, E. G. *Nature* **320**, 750–756 (1986).

doi:10.1038/nature04360

Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*

Yun Zhang¹, Hang Lu¹ & Cornelia I. Bargmann¹

Food can be hazardous, either through toxicity or through bacterial infections that follow the ingestion of a tainted food source. Because learning about food quality enhances survival, one of the most robust forms of olfactory learning is conditioned avoidance of tastes associated with visceral malaise. The nematode *Caenorhabditis elegans* feeds on bacteria but is susceptible to infection by pathogenic bacteria in its natural environment. Here we show that *C. elegans* modifies its olfactory preferences after exposure to pathogenic bacteria, avoiding odours from the pathogen and increasing its attraction to odours from familiar nonpathogenic bacteria. Particular bacteria elicit specific changes in olfactory preferences that are suggestive of associative learning. Exposure to pathogenic bacteria increases serotonin in ADF chemosensory neurons by transcriptional and post-transcriptional mechanisms. Serotonin functions through MOD-1, a serotonin-gated chloride channel expressed in sensory interneurons, to promote aversive learning. An increase in serotonin may represent the negative reinforcing stimulus in pathogenic infection.

Many animals are susceptible to intestinal infections by bacteria. The pathogenic soil bacteria *Pseudomonas aeruginosa* and *Serratia marcescens* can proliferate in the intestine of the soil nematode *C. elegans* after they are ingested, resulting in death of the nematode after several days^{1–3}. *C. elegans* protects itself from pathogens through innate immunity pathways^{3,4} and through behavioural strategies such as leaving a lawn of pathogenic bacteria². *C. elegans* has a simple nervous system of 302 neurons that facilitates the identification of molecules, neurons and circuits involved in behaviour⁵. One of its most robust behaviours is olfactory chemotaxis towards food-associated odours, an innate behaviour that is highly reproducible among animals⁶. Olfactory preference can be altered by adaptation after prolonged exposure to an odour^{7,8} or by starvation^{9,10}. Here we use infection by natural pathogens to develop an ecologically relevant olfactory learning assay, with which we identify a circuit and neuronal changes that are associated with learning.

Pathogenic bacteria alter odour preference

The preference of *C. elegans* for different bacterial odours can be measured in a binary choice assay in which animals migrate towards one of two bacterial lawns on opposite sides of a plate (Fig. 1a). In this assay, a choice index of -1.0 represents complete preference for *Escherichia coli* OP50, the control bacterium in all tests, an index of 1.0 represents complete preference for the test bacterium, and an index of 0 represents an equal distribution (Fig. 1a). Animals cultivated on OP50 alone were equally attracted to OP50 and the pathogenic *P. aeruginosa* strain PA14, and were more attracted to the pathogenic bacterium *S. marcescens* ATCC 13880 than to OP50, despite the eventual toxicity of *S. marcescens* infection (Fig. 1b and Supplementary Fig. 1). Animals cultivated from hatching in the presence of both OP50 and PA14, however, strongly preferred OP50 to *P. aeruginosa* PA14; similarly, animals cultivated on OP50 and *S. marcescens* preferred OP50 to *S. marcescens* (Fig. 1b). It was not

possible to raise animals on *S. marcescens* or PA14 alone because of the virulence of the infection¹. These results indicate that *C. elegans* can modify its olfactory preferences to avoid toxic bacteria.

No alteration in olfactory preferences was observed when animals were raised on both OP50 and nonpathogenic *E. coli* HB101, or on OP50 and the harmless soil bacteria *Rhizobium leguminosarum* or *Pseudomonas fluorescens* (Fig. 1b). Three isogenic nonvirulent derivatives of PA14, 12A1, 50E12 and PA14 (*gacA::Kan*)¹, and two other nonpathogenic strains of *P. aeruginosa*, PAK and PA103, did not modify the olfactory preferences of *C. elegans* (Fig. 1b). These results suggest that pathogenic infection induces the alteration in olfactory preferences.

A learning index was generated by subtracting the choice index of animals exposed to pathogen from the choice index of naive animals. A positive learning index, as shown by wild-type animals exposed to PA14 or *S. marcescens* (Fig. 1c), indicates an acquired avoidance of pathogenic bacteria. When adult animals were acutely exposed to PA14 for only 4 h, the learning index was similar to that of animals that had had lifelong exposure to OP50 and PA14 (Fig. 1d). Adult animals that were grown on *E. coli* OP50 and then starved for 4 h did not show the same change in olfactory preference (Fig. 1e). This result indicates that adult animals rapidly modify their olfactory preferences after exposure to pathogenic bacteria.

Olfactory learning in response to pathogenic bacteria was distinct from known forms of odour adaptation, because the olfactory adaptation-defective mutants *egl-4(ky95)* and *adp-1(ky20)*^{7,11} were both proficient in olfactory preference learning after exposure to OP50 and PA14 (Supplementary Figs 2 and 3). Olfactory learning is also distinct from the food-leaving behaviour that *C. elegans* shows in response to *S. marcescens* Db10 or Db11 (ref. 2). Food-leaving requires the Toll-like receptor TOL-1, but *tol-1(nr2033)* animals were proficient in olfactory preference learning (Supplementary Figs 2 and 3). These results suggest that *C. elegans* uses several different behavioural strategies to minimize exposure to pathogens.

¹Howard Hughes Medical Institute, Laboratory of Neural Circuits and Behavior, The Rockefeller University, New York, New York 10021, USA.

To test whether *C. elegans* associates pathogenesis with simultaneously presented odours, we exposed animals sequentially to pathogenic or nonpathogenic bacterial strains. In the first experiment, adult animals were exposed to pathogenic *P. aeruginosa* PA14 for 4 h, followed by nonpathogenic *S. marcescens* Db1140 for 4 h. In the second experiment, adults were exposed to nonpathogenic *P. aeruginosa* 50E12 for 4 h followed by pathogenic *S. marcescens* ATCC 13880 for 4 h. In the third and fourth experiments, the order of presentation for pathogenic and nonpathogenic strains was reversed. Trained animals were then tested for their preference between PA14 and *S. marcescens*. In all tests, animals exposed to pathogenic PA14 and harmless *S. marcescens* showed enhanced avoidance of PA14 as compared with those exposed to pathogenic *S. marcescens* and harmless *P. aeruginosa* (Fig. 1f). This dissociation experiment indicates that *C. elegans* selectively avoids an odour experienced at the same time as pathogenic infection, a criterion for associative learning.

Aversive and attractive aspects of learning

The two-choice preference assay does not distinguish whether trained animals have made a positive association that increases attraction towards the harmless bacterium or a negative association that induces aversion from the pathogenic bacterium. We tested these possibilities by a four-choice maze assay in which animals trained with OP50 and *P. aeruginosa* PA14 were given a choice between OP50, PA14 and two novel bacteria strains: one nonpathogenic strain, *P. fluorescens*; and one pathogenic strain, *S. marcescens* ATCC 13880. Preferences were tested in a partially enclosed eight-

arm maze made from polydimethyl siloxane (PDMS) elastomer resting on the surface of an agar plate (Fig. 2a). A decision area in the centre of the maze was connected by slender channels to eight small food chambers, each containing one of the four bacterial strains. Animals were placed in the open decision area and approached food chambers through the channels. Wild-type animals cultivated on OP50 distributed themselves among all four bacteria strains, reproducibly showing strongest attraction towards *S. marcescens* (Fig. 2b).

The four-choice configuration made it possible to compare responses to the two bacteria experienced during training with responses to the novel bacteria. After cultivation on OP50 and PA14, the fraction of animals that approached OP50 was increased, and the fraction of animals that approached PA14 was diminished, as compared with control bacteria (Fig. 2b, c, and Methods), suggesting that olfactory learning on pathogens includes both attractive and aversive components. Similarly, animals cultivated on a different pairing of nonpathogenic and pathogenic bacteria, *P. fluorescens* and *S. marcescens* ATCC 13880, showed increased attraction to *P. fluorescens* and aversion from *S. marcescens* in the four-choice maze assay (Fig. 2d, e).

Exposure to PA14 for 4 h was sufficient to induce aversion from PA14 but not increased attraction towards OP50 (Supplementary Fig. 4). These results suggest that the aversive component of olfactory learning is relatively rapid, whereas attractive changes occur more slowly. Starvation does not elicit these changes in preference (Supplementary Fig. 4).

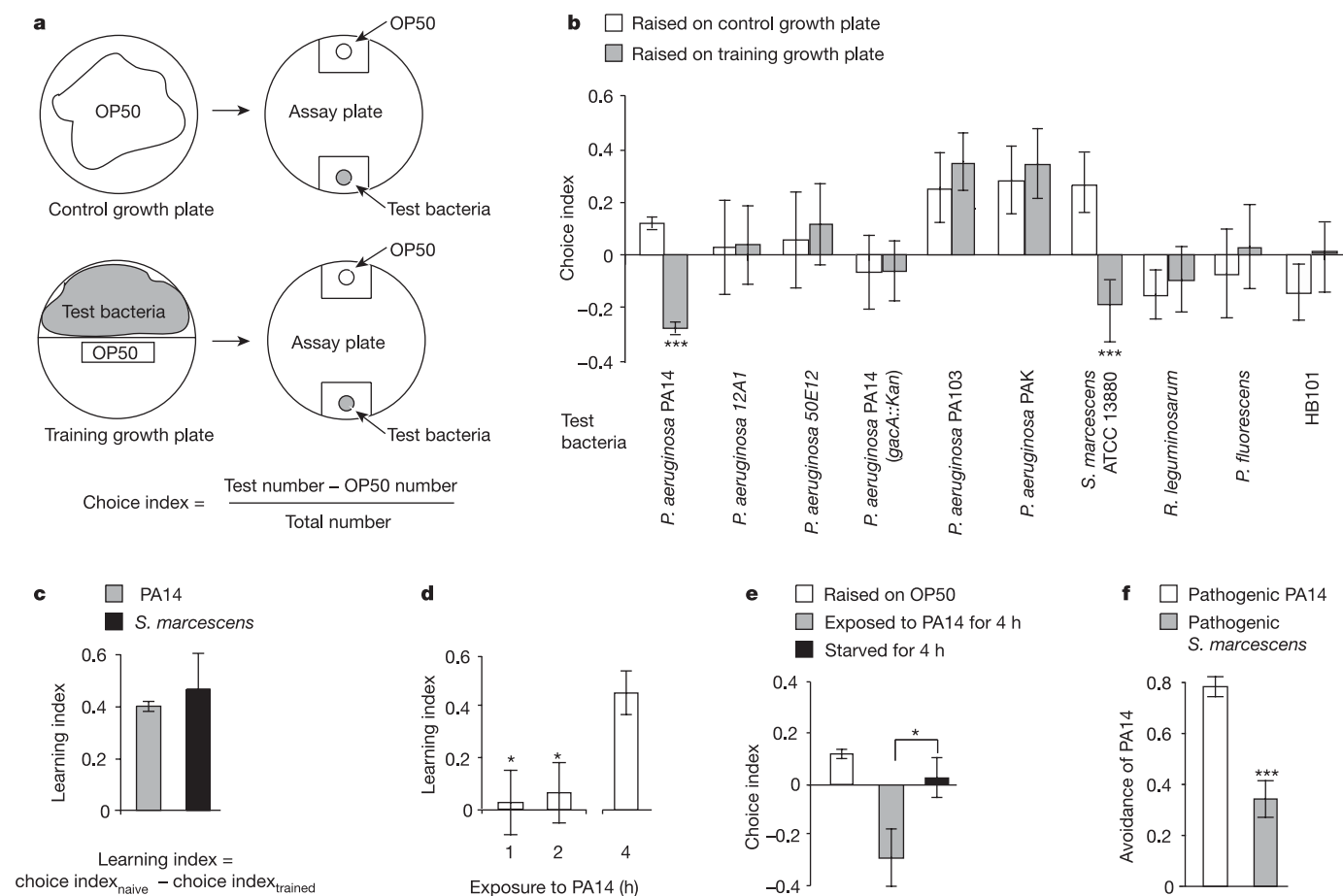


Figure 1 | *C. elegans* learns to avoid pathogenic bacteria. **a**, Training protocol and two-choice olfactory preference assays. **b**, Olfactory preferences after exposure to pathogenic *P. aeruginosa* PA14 or *S. marcescens*, isogenic nonvirulent *P. aeruginosa*, or nonpathogenic bacteria. **c**, Learning index after training with pathogens. **d**, Rapid learning in adults

transferred from OP50 to PA14. **e**, Starvation and pathogens cause different changes in olfactory preference. **f**, Animals exposed to pathogenic *P. aeruginosa* PA14 and harmless *S. marcescens* avoid PA14 more than do animals exposed to harmless *P. aeruginosa* and pathogenic *S. marcescens*. *** $P < 0.001$, * $P < 0.05$, $n \geq 4$ assays. Error bars indicate the s.e.m.

Serotonin induces aversive learning

The neurotransmitter serotonin is essential for pathogen-induced olfactory learning. The general catecholamine-defective mutants *cat-1* (ref. 12) and *cat-4* (ref. 13) showed significantly reduced learning in the two-choice assay (Fig. 3a and Supplementary Fig. 3), as did the more specific mutant *tph-1*, which is deficient in a tryptophan hydroxylase required for biosynthesis of serotonin but not other catecholamines¹⁴ (Fig. 3a and Supplementary Fig. 3). *cat-2* mutants lacking dopamine¹⁵ but not serotonin were proficient in learning. In the four-choice maze assay, *tph-1* mutants were defective in both aversive and attractive components of olfactory learning (Fig. 3b and Supplementary Fig. 9a, b).

tph-1 mutants were normal in their basal preference for bacterial strains and in their tendency to leave a lawn of pathogenic *S. marcescens* (Supplementary Figs 5 and 9a) and showed no alteration in their susceptibility to pathogenic infection and killing by PA14 (Supplementary Fig. 6). Thus, the learning defects of *tph-1* mutants are not due to changes in general olfactory ability, recognition of bacteria or innate immunity. Instead, *tph-1* is selectively unable to associate the physiological responses to pathogens with olfactory cues.

In *C. elegans* hermaphrodites, *tph-1* is expressed in the serotonergic neurons ADF, NSM and HSN, and occasionally AIM and RIH¹⁴. Expression of a *tph-1* complementary DNA in ADF chemosensory neurons partially rescued the learning defects of *tph-1* mutants in the two-choice assay, but expression of *tph-1* in NSM pharyngeal neurons did not (Fig. 3c and Supplementary Fig. 3). Expression of *tph-1* in NSM neurons did partially rescue a different serotonin-dependent behaviour, namely the enhanced slowing of starved animals in response to fresh food, but this behaviour was not rescued by *tph-1* expression in ADF neurons¹⁶ (Supplementary Fig. 7). These results suggest that serotonin from ADF and NSM neurons has

different functions: serotonin from ADF neurons has a stronger role in olfactory learning, whereas that from NSM neurons has a stronger role in the enhanced slowing response.

In the four-choice maze assay, expression of *tph-1* in ADF neurons alone fully rescued aversive but not attractive learning (Fig. 3d and Supplementary Fig. 9d, e). Expression of *tph-1* in NSM neurons did not rescue learning, but expression in both ADF and NSM neurons restored both aversive and attractive learning (Fig. 3d and Supplementary Fig. 9c). We propose that ADF neurons may evaluate aversive and NSM neurons attractive components of food-related signals.

The *C. elegans* genome encodes at least 12 potential serotonin receptors, including MOD-1, a serotonin-gated chloride channel that regulates the enhanced slowing of starved animals on fresh food^{16,17}. *mod-1* mutants showed significantly decreased olfactory learning to PA14 in the two-choice learning assay (Fig. 4a and Supplementary Fig. 3). In the four-choice maze assay, *mod-1* mutants were specifically defective in aversive but not attractive learning when trained either on OP50 and PA14, or on *P. fluorescens* and *S. marcescens* (Fig. 4b and Supplementary Fig. 10a, b). Thus, the partial defect in *mod-1* seems to result from its role in aversive learning, the same component that is affected by ADF sensory neurons.

Like *tph-1* mutants, *mod-1* mutants were killed by PA14 infection with the same kinetics as wild-type animals and were able to discriminate between bacteria in maze assays (Supplementary Figs 6 and 10a). The specific defect of *mod-1* mutants in aversive learning suggests that MOD-1 is the downstream target of the serotonin signal from ADF neurons.

mod-1 promoter fusions are expressed in AIA, AIB, AIY, RID and probably AIZ interneurons, as well as in other neurons in the head, ventral cord and tail^{17,18}. The aversive learning defect of *mod-1(ok103)* mutants was completely rescued by expression of a

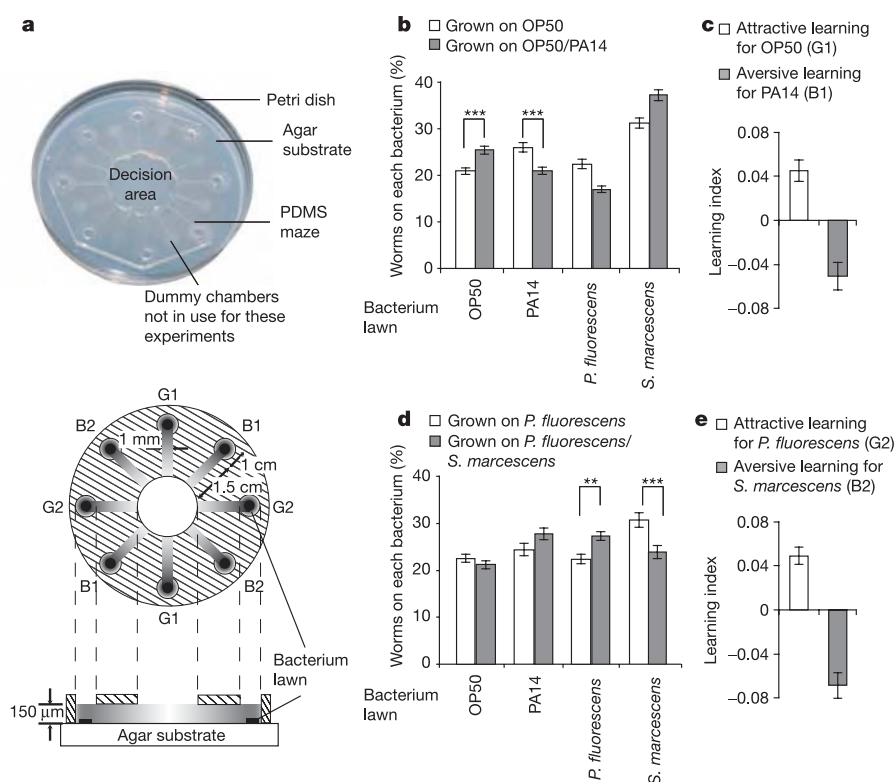


Figure 2 | Olfactory maze assay. **a**, Photograph and scheme of the four-choice maze assay for olfactory preference. Each test bacterium was placed in two of the eight chambers (G1, *E. coli* OP50; B1, *P. aeruginosa* PA14; G2, *P. fluorescens*; B2, *S. marcescens* ATCC 13880). **b**, **c**, Training with PA14 and OP50 results in both increased attraction towards OP50 and

aversion from PA14. **d**, **e**, Training with *S. marcescens* and *P. fluorescens* results in both increased attraction towards *P. fluorescens* and aversion from *S. marcescens*. *** $P < 0.001$, ** $P < 0.01$, $n \geq 23$ assays. Error bars indicate the s.e.m.

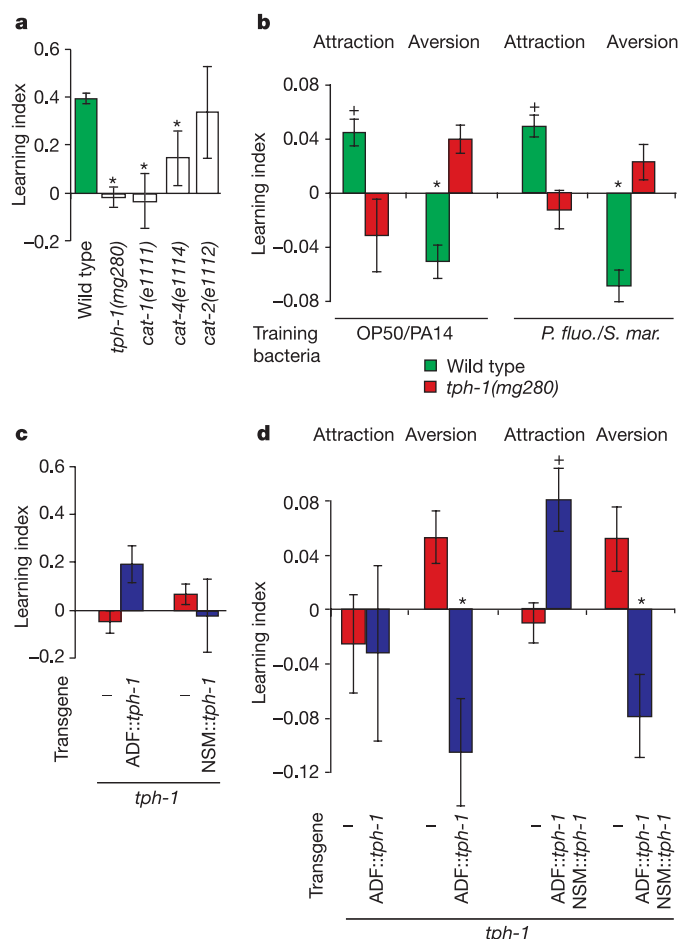


Figure 3 | ADF serotonergic neurons regulate aversive learning. **a**, Two-choice learning to PA14 in catecholamine biosynthesis mutants. **b**, Defective aversive and attractive learning in four-choice maze assays of *tph-1(mg280)* animals trained with PA14 or *S. marcescens*. **c**, *tph-1* expression in ADF neurons partially rescues *tph-1(mg280)* in two-choice assays. Expression in NSM and other pharyngeal neurons does not rescue. **d**, *tph-1* expression in ADF neurons in *tph-1(mg280)* rescues aversive learning to PA14, but not attractive learning to OP50. *tph-1* expression in both ADF and NSM neurons rescues both attractive and aversive learning. * $P < 0.05$, + $P < 0.05$, $n \geq 6$ assays. Error bars indicate the s.e.m.

mod-1 cDNA from a *mod-1* promoter¹⁸ (Fig. 4c and Supplementary Fig. 10c, d). The principal interneurons downstream of chemosensory neurons such as ADF are AIA, AIB, AIY and AIZ⁵. We used a *ttx-3* promoter to express *mod-1* in AIY and possibly AIA interneurons¹⁸, and an *odr-2(2b)* promoter to express wild-type *mod-1* in AIZ and AIB interneurons, as well as in a few other neurons¹⁹. Both the *ttx-3::mod-1* and *odr-2(2b)::mod-1* transgenes rescued aversive learning in *mod-1(ok103)* mutants (Fig. 4c and Supplementary Fig. 10e–h), suggesting that the serotonin receptor can function in several interneurons to modulate olfactory preference. Because ADF neurons synapse onto AIZ and perhaps AIY interneurons⁵, these two neurons are potential sites of MOD-1 action in aversive learning.

To determine how serotonin signalling changes during learning, we examined serotonin immunoreactivity in OP50-fed naive animals and PA14-trained animals by staining with polyclonal antibodies against serotonin. Exposure to PA14 resulted in a 3.3 ± 0.3 (s.e.m.) fold increase in serotonin immunostaining in ADF neurons, but no change in NSM neurons (Fig. 5a, b, i). A significant increase in serotonin in PA14-trained animals as compared with OP50-fed animals was also detected by directly measuring serotonin in dialysed *C. elegans* homogenates by high-performance liquid chromatography (HPLC; data not shown). No serotonin was detected by

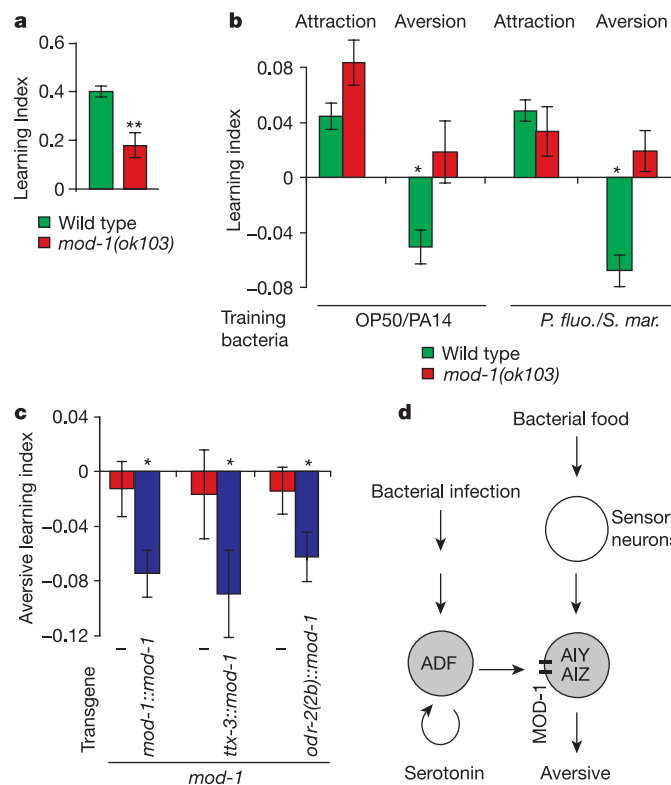


Figure 4 | MOD-1 regulates aversive learning. **a**, Two-choice learning assays to PA14. *mod-1(ok103)* is partly defective. **b**, Four-choice maze assays. *mod-1(ok103)* animals are defective in aversive learning when trained either with OP50 and PA14 or with *P. fluorescens* and *S. marcescens*. **c**, Four-choice maze assays for aversive learning after training with OP50 and PA14. Expression of *mod-1* in several subsets of *mod-1*-expressing neurons rescues aversive learning. *ttx-3::mod-1* is expressed in AIY neurons; *odr-2(2b)::mod-1* is expressed in AIB, AIZ and other neurons. **d**, Model of aversive olfactory learning on pathogenic bacteria. ** $P < 0.01$, * $P < 0.05$, $n \geq 5$ assays. Error bars indicate the s.e.m.

HPLC or antibody staining in PA14-trained *tph-1* mutants. Similarly, animals exposed to *S. marcescens* showed a 2.2 ± 0.2 fold increase in ADF immunoreactivity as compared with animals fed on nonpathogenic *P. fluorescens* (Fig. 5c, d, i). Serotonin immunoreactivity in ADF derivatives that did not induce olfactory learning (Supplementary Fig. 8). These results suggest that exposure to pathogenic bacteria specifically increases serotonin in ADF neurons.

Serotonin in ADF neurons could rise either through increased transcription of *tph-1*, the rate-limiting enzyme for serotonin biosynthesis, or through post-transcriptional mechanisms such as changes in TPH-1 enzymatic activity. The transcription of *C. elegans tph-1* in ADF neurons is increased by neuronal activity²⁰, recovery from the dauer stage, and heat stress²¹. The biochemical activity of mammalian tryptophan hydroxylase is activated by phosphorylation mediated by Ca^{2+} /calmodulin-mediated kinase II and protein kinase A (ref. 22). To characterize the mechanism of serotonin upregulation in ADF neurons, we examined serotonin immunoreactivity in a *tph-1;srh-142::tph-1* strain in which TPH-1 was expressed from a heterologous, ADF-specific promoter. The *srh-142* promoter was not regulated by exposure to PA14 (data not shown). In this strain, exposure to PA14 caused a 2.1 ± 0.15 fold increase in serotonin immunoreactivity in ADF as compared with naive OP50-fed animals (Fig. 5e, f, j). Serotonin immunoreactivity in NSM neurons was also detected in these animals, probably owing to the reuptake of serotonin released from ADF neurons. These results suggest that pathogen exposure increases TPH-1 activity or decreases serotonin turnover in

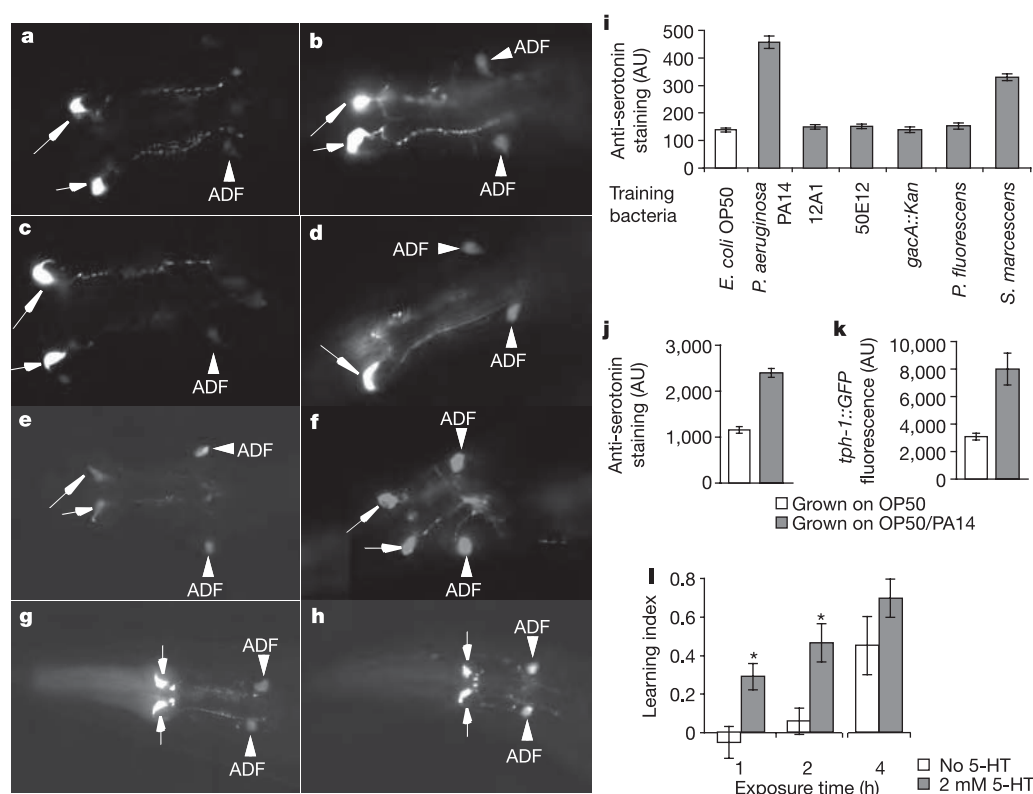


Figure 5 | Pathogenic bacteria increase serotonin in ADF neurons. **a–f**, Serotonin immunoreactivity in either wild-type animals fed OP50 (**a**), OP50 and PA14 (**b**), *P. fluorescens* (**c**) or *P. fluorescens* and *S. marcescens* (**d**), or *tph-1::srh-142::tph-1* animals fed OP50 (**e**) or OP50 and PA14 (**f**). Arrowheads indicate ADF neurons; arrows indicate NSM neurons. **g, h**, *tph-1::GFP* expression in wild-type animals fed OP50 (**g**), or OP50 and PA14 (**h**). **i**, ADF serotonin immunoreactivity in wild-type animals fed OP50 with other bacteria. **j**, ADF serotonin immunoreactivity in *tph-1::srh-142::tph-1* animals. **k**, ADF *tph-1::GFP* fluorescence in wild-type animals. **l**, Exogenous serotonin (5-HT) accelerates olfactory learning. * $P < 0.05$, $n \geq 4$ assays or ≥ 14 animals. Error bars indicate the s.e.m. AU, arbitrary fluorescence units.

ADF neurons. The rescue of aversive learning by *srh-142::tph-1* indicates that post-transcriptional mechanisms are sufficient for learning (Fig. 3d). Notably, exposure to PA14 also induced a 2.6 ± 0.42 fold increase in expression of a *tph-1::GFP* reporter gene lacking most of the TPH-1 protein in ADF neurons (Fig. 5g, h, k). Thus, pathogens increase serotonin in ADF neurons by both transcriptional and post-transcriptional mechanisms.

We next tested whether increased serotonin facilitates olfactory learning directly. Taking advantage of the ability of *C. elegans* to take up exogenous serotonin¹⁶, we raised wild-type animals on OP50 and transferred them to PA14 with or without 2 mM exogenous serotonin. In the presence of exogenous serotonin, significant learning was observed within 1 h of pathogen exposure, and full aversive learning within 2 h (as compared with 4 h in untreated animals; Figs 1d and 5l). These results suggest that an increase in serotonin directly promotes olfactory learning in pathogen-exposed animals, perhaps by encoding the unconditioned stimulus of pathogenic infection.

Discussion

In its natural soil habitat, *C. elegans* interacts with many different bacteria. Some are good food sources, some are poor food sources, and some are pathogenic hazards^{1–3,23}. Here we have shown that *C. elegans* learns to avoid the odours of pathogenic bacteria after interacting with the pathogens. Exposure to pathogens upregulates expression of serotonin in the ADF chemosensory neurons, and aversive learning requires serotonin from ADF neurons and the serotonin receptor MOD-1 (Fig. 4d). The induced avoidance of pathogenic bacteria is analogous to conditioned taste aversion, a learning behaviour that has been described in mammals, snails, cuttlefish and fish^{24–26} in which animals avoid food flavours associated with intestinal distress. Olfactory learning may allow *C. elegans* to distinguish among natural food sources on the basis of relevant experiences.

In both vertebrates and invertebrates, catecholamines including serotonin function as reinforcing signals during learning²⁷. In mammals, dopamine and norepinephrine are reinforcing signals in

learning and addiction²⁸. In insect olfactory learning, octopamine functions as a positive reinforcing signal and dopamine as a negative reinforcing signal²⁹. Serotonin, particularly that from NSM neurons, has been considered to be a positive, food-related signal in *C. elegans*: serotonin falls when *C. elegans* is removed from food, and its absence is associated with starvation-induced behaviours^{9,10,14,16,30}. Our results identify another role of serotonin: in ADF neurons, serotonin is increased under noxious conditions of infection by pathogenic bacteria. The rapid transcriptional and post-transcriptional induction of serotonin after pathogen exposure provides a signal that could be used to modify various behaviours.

Serotonin in the mammalian intestine functions in signalling malaise, specifically the nausea associated with chemotherapy, by activating the 5-HT₃ receptor, a serotonin-gated ion channel³¹. Our results indicate that MOD-1, a serotonin-gated ion channel of *C. elegans*, signals the presence of aversive intestinal pathogens. The similarity of these vertebrate and invertebrate pathways might result from an ancient role of serotonin in signalling between the viscera and the brain; 95% of the serotonin in the human body is made by intestinal cells, not neurons³¹.

METHODS

Nematode strains, molecular biological methods, immunohistochemistry and statistics are described in the Supplementary Information.

Binary choice assays. Embryos were collected by bleaching and were grown at room temperature. 'Naive' animals were grown on a standard nematode growth medium (NGM) plate that was evenly spread with 300 μ l of *E. coli* OP50 suspension. For training, a suspension of ~ 200 μ l of the test bacteria was spread on a plate and ~ 50 μ l of OP50 suspension was used to make a small lawn on the side. Plates were incubated at 26 °C for 48 h before use.

The two-choice olfactory preference assays were based on standard chemotaxis assays⁶ except that bacterial suspensions were used as odour sources (Fig. 1a). Bacteria grown overnight in NGM at 26 °C were resuspended at an absorbance of 1.0 at 600 nm, and 25 μ l of each bacterial suspension was spotted onto the plate and air-dried for 5 h at room temperature. Animals were washed twice in S-basal buffer and once in assay buffer, and 50–200 animals were placed near the centre of the plate, equidistant from the two bacteria. Animals were allowed to move freely for 1–2 h before being immobilized by 1 μ l of 10 mM

sodium azide applied at the bacteria spots. In most cases, animals quickly entered one lawn and remained there for the duration of the assay.

Microfabrication and four-choice maze assays. We fabricated microdevices using the PDMS rapid prototyping technique³². Photolithography masks were laser-printed on emulsion films with 5,080 d.p.i. resolution and used to produce prototype masters in a photo-patternable epoxy resin (SU-8-50, Microchem) on silicon wafers by ultraviolet photolithography. Masters were silanized by vapour-phase tridecafluoro-1,1,2,2-tetrahydrooctyl trichlorosilane (United Chemical Technologies). PDMS devices were micro-moulded using two-part Sylgard 184 silicone elastomer (Dow Corning). Small holes were punched out above the decision area and microwells for loading *C. elegans* and bacteria. A maze was placed on an assay plate immediately before bacteria suspensions were spotted onto the plate. Bacteria were prepared as in the two-choice assays, except that suspensions were 10 times as concentrated for *S. marcescens* and 20 times as concentrated for the other bacteria. We used 1.25 µl of bacteria suspension in each chamber.

An attractive learning index for OP50 (G1) was calculated as (percentage of animals at G1)_{trained} – (percentage of animals at G1)_{naive}. An aversive learning index for PA14 (B1) was calculated as (percentage of animals at B1)_{trained} – (percentage of animals at B1)_{naive}. An attractive learning index for *P. fluorescens* (G2) and an aversive learning index for *S. marcescens* (B2) were calculated in the same way as described for the OP50 and PA14 training.

Received 13 July; accepted 8 September 2005.

1. Tan, M. W., Rahme, L. G., Sternberg, J. A., Tompkins, R. G. & Ausubel, F. M. *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors. *Proc. Natl Acad. Sci. USA* **96**, 2408–2413 (1999).
2. Pujol, N. et al. A reverse genetic analysis of components of the Toll signalling pathway in *Caenorhabditis elegans*. *Curr. Biol.* **11**, 809–821 (2001).
3. Nicholas, H. R. & Hodgkin, J. Responses to infection and possible recognition strategies in the innate immune system of *Caenorhabditis elegans*. *Mol. Immunol.* **41**, 479–493 (2004).
4. Kurz, C. L. & Ewbank, J. J. *Caenorhabditis elegans*: an emerging genetic model for the study of innate immunity. *Nature Rev. Genet.* **4**, 380–390 (2003).
5. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
6. Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).
7. Colbert, H. A. & Bargmann, C. I. Odorant-specific adaptation pathways generate olfactory plasticity in *C. elegans*. *Neuron* **14**, 803–812 (1995).
8. Bernhard, N. & van der Kooy, D. A behavioural and genetic dissection of two forms of olfactory plasticity in *Caenorhabditis elegans*: adaptation and habituation. *Learn. Mem.* **7**, 199–212 (2000).
9. Colbert, H. A. & Bargmann, C. I. Environmental signals modulate olfactory acuity, discrimination, and memory in *Caenorhabditis elegans*. *Learn. Mem.* **4**, 179–191 (1997).
10. Chao, M. Y., Komatsu, H., Fukuto, H. S., Dionne, H. M. & Hart, A. C. Feeding status and serotonin rapidly and reversibly modulate a *Caenorhabditis elegans* chemosensory circuit. *Proc. Natl Acad. Sci. USA* **101**, 15512–15517 (2004).
11. L'Etoile, N. D. et al. The cyclic GMP-dependent protein kinase EGL-4 regulates olfactory adaptation in *C. elegans*. *Neuron* **36**, 1079–1089 (2002).
12. Duerr, J. S. et al. The *cat-1* gene of *Caenorhabditis elegans* encodes a vesicular monoamine transporter required for specific monoamine-dependent behaviors. *J. Neurosci.* **19**, 72–84 (1999).
13. Weinshenker, D., Garriga, G. & Thomas, J. H. Genetic and pharmacological analysis of neurotransmitters controlling egg laying in *C. elegans*. *J. Neurosci.* **15**, 6975–6985 (1995).
14. Sze, J. Y., Victor, M., Loer, C., Shi, Y. & Ruvkun, G. Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. *Nature* **403**, 560–564 (2000).
15. Lints, R. & Emmons, S. W. Patterning of dopaminergic neurotransmitter identity among *Caenorhabditis elegans* ray sensory neurons by a TGFβ family signalling pathway and a *Hox* gene. *Development* **126**, 5819–5831 (1999).
16. Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
17. Ranganathan, R., Cannon, S. C. & Horvitz, H. R. MOD-1 is a serotonin-gated chloride channel that modulates locomotory behaviour in *C. elegans*. *Nature* **408**, 470–475 (2000).
18. Wenick, A. S. & Hobert, O. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev Cell* **6**, 757–770 (2004).
19. Chou, J. H., Bargmann, C. I. & Sengupta, P. The *Caenorhabditis elegans odr-2* gene encodes a novel Ly-6-related protein required for olfaction. *Genetics* **157**, 211–224 (2001).
20. Zhang, S., Sokolchik, I., Blanco, G. & Sze, J. Y. *Caenorhabditis elegans* TRPV ion channel regulates 5HT biosynthesis in chemosensory neurons. *Development* **131**, 1629–1638 (2004).
21. Estevez, M., Estevez, A. O., Cowie, R. H. & Gardner, K. L. The voltage-gated calcium channel UNC-2 is involved in stress-mediated regulation of tryptophan hydroxylase. *J. Neurochem.* **88**, 102–113 (2004).
22. Mockus, S. M. & Vrana, K. E. Advances in the molecular characterization of tryptophan hydroxylase. *J. Mol. Neurosci.* **10**, 163–179 (1998).
23. Avery, L. & Shtonda, B. B. Food transport in the *C. elegans* pharynx. *J. Exp. Biol.* **206**, 2441–2457 (2003).
24. Bernstein, I. L. Taste aversion learning: a contemporary perspective. *Nutrition* **15**, 229–234 (1999).
25. Carew, T. J. & Sahley, C. L. Invertebrate learning and memory: from behaviour to molecules. *Annu. Rev. Neurosci.* **9**, 435–487 (1986).
26. Darmilaq, A., Dickel, L., Chichery, M., Agin, V. & Chichery, R. Rapid taste aversion learning in adult cuttlefish, *Sepia officinalis*. *Anim. Behav.* **68**, 1291–1298 (2004).
27. Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
28. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
29. Schwaerzel, M. et al. Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in *Drosophila*. *J. Neurosci.* **23**, 10495–10502 (2003).
30. Nuttley, W. M., Atkinson-Leadbetter, K. P. & Van Der Kooy, D. Serotonin mediates food-odor associative learning in the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **99**, 12449–12454 (2002).
31. Gershon, M. D. Plasticity in serotonin control mechanisms in the gut. *Curr. Opin. Pharmacol.* **3**, 600–607 (2003).
32. Duffy, D. C., McDonald, J. C., Schueller, O. J. A. & Whitesides, G. M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Analytical Chemistry* **70**, 4974–4984 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. N. Engel for *P. aeruginosa* strains PA14, PAK and PA103; J. Y. Sze for the *tph-1::GFP* strain; L. Tecott and O. Murphy for assistance with HPLC; L. G. Rahme and F. M. Ausubel for *P. aeruginosa* strains 50E12, 12A1 and PA14 (*gacA::Kan*); and J. J. Ewbank for *S. marcescens* strains Db11 and Db1140 and for comments on the manuscript. Y.Z. was supported by an American Heart Association Postdoctoral Fellowship and the Howard Hughes Medical Institute. H.L. was supported by a Mentored Quantitative Scientist Career Development Award from the NIH and the Howard Hughes Medical Institute. C.I.B. is an investigator of the Howard Hughes Medical Institute. This work was supported by funding from the Howard Hughes Medical Institute (to C.I.B.).

Author Contributions Y.Z. and H.L. performed all experiments. C.I.B., Y.Z. and H.L. contributed to experimental design and analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.I.B. (cori@rockefeller.edu).

Subunit arrangement and function in NMDA receptors

Hiroyasu Furukawa^{1†}, Satinder K Singh^{1†}, Romina Mancusso^{1†} & Eric Gouaux^{1,2†}

Excitatory neurotransmission mediated by NMDA (*N*-methyl-D-aspartate) receptors is fundamental to the physiology of the mammalian central nervous system. These receptors are heteromeric ion channels that for activation require binding of glycine and glutamate to the NR1 and NR2 subunits, respectively. NMDA receptor function is characterized by slow channel opening and deactivation, and the resulting influx of cations initiates signal transduction cascades that are crucial to higher functions including learning and memory. Here we report crystal structures of the ligand-binding core of NR2A with glutamate and that of the NR1–NR2A heterodimer with glutamate and glycine. The NR2A–glutamate complex defines the determinants of glutamate and NMDA recognition, and the NR1–NR2A heterodimer suggests a mechanism for ligand-induced ion channel opening. Analysis of the heterodimer interface, together with biochemical and electrophysiological experiments, confirms that the NR1–NR2A heterodimer is the functional unit in tetrameric NMDA receptors and that tyrosine 535 of NR1, located in the subunit interface, modulates the rate of ion channel deactivation.

Glutamate, a simple amino acid, is an essential currency of the human nervous system and is transmitted from one neuron to another at specialized junctions called synapses. The tightly regulated release of glutamate from one neuron, coupled with its detection by glutamate receptors on the adjacent neuron, forms the basis of synaptic transmission at many of the $\sim 10^{14}$ synapses in the human brain¹. Specificity of synaptic signalling by glutamate in space and time is conferred by the precise positioning of synapses and by the neuron-specific expression of a subset of genes encoding glutamate receptors. Pharmacological studies provided initial clues to the diversity of glutamate receptor proteins, and early studies partitioned them into two classes depending on their response to the synthetic agonist NMDA². Subsequent cloning of glutamate receptor genes and analysis of their predicted protein sequences facilitated the clustering of NMDA and non-NMDA receptors into distinct protein families^{3–7}.

NMDA receptors are unusual ligand-gated ion channels because activation not only requires the binding of two agonists, glycine and glutamate⁸, but also demands the relief of Mg^{2+} block by membrane depolarization⁹. The opening of NMDA receptors leads to an influx of cations including Ca^{2+} , and the permeation of Ca^{2+} through NMDA receptor ion channels¹⁰ initiates signal transduction cascades that in turn modulate synaptic strength. The rates at which the responses of NMDA receptors rise (activate) and decline (deactivate) upon application and removal of agonists, respectively, are markedly slower than those of non-NMDA receptors^{11,12} and it is the slow deactivation rate of NMDA receptors that governs the duration of the excitatory postsynaptic potential, a measure of the 'strength' of synaptic signalling¹³. The integration of chemical and electrical stimuli by NMDA receptors into a Ca^{2+} signal is crucial for activity-dependent synaptic plasticity, which in turn underpins many higher functions including learning and memory. By contrast, dysfunction of NMDA receptors has been implicated in many diseases and injuries including stroke, Parkinson's disease, Huntington's disease and schizophrenia^{14,15}.

The functional diversity shown by NMDA receptors is rooted in their assembly as obligate heteromers of glycine-binding NR1, glutamate-binding NR2 and glycine-binding NR3 subunits. Whereas non-NMDA receptors such as AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and kainate receptors can form functional homotetrameric channels activated solely by glutamate, NMDA receptors require the assembly of two copies each of the NR1 and NR2 and/or NR3 subunits^{16,17}. Contingent on the cell and the developmental stage, typically one of the four NR2 subunits (A–D) combines with a splice variant of the NR1 subunit, yielding receptors with distinct deactivation kinetics^{18,19}. The apparent affinity of the NR1 subunit for glycine depends on the identity of the coassembled NR2 subunit, which suggests that there is allosteric coupling between NR1 and NR2²⁰. Moreover, for a particular combination of NR1 and NR2 subunits, glycine and glutamate binding show negative cooperativity, giving rise to a glycine-dependent form of receptor desensitization²¹.

Studies of AMPA receptors have shown that the receptor subunits assemble as a dimer of dimers through interactions between different domains on each subunit, including the amino terminal domain (ATD) and the ligand-binding domain (S1S2)²² (Fig. 1a), and dimers of the ligand-binding domain of the GluR2 receptor have been particularly well studied^{23,24}. Although it has been speculated that NMDA receptors are also organized as a dimer of dimers with an NR1–NR1–NR2–NR2 arrangement²⁵ mediated by the ATD²⁶ and part of the S1 segment²⁷, conclusive proof of the subunit arrangement and the nature of subunit–subunit contacts is lacking (Fig. 1b). In addition, mechanistic understanding of the role that subunit–subunit contacts might have in NMDA receptor activity is absent.

Here we present crystal structures of the ligand-binding core of NR2A bound to glutamate and that of the NR1–NR2A heterodimer bound to glycine and glutamate. The physiological relevance of the subunit arrangement observed in the NR1–NR2A crystal structure is confirmed by biochemical and electrophysiological experiments. Significantly, detailed analysis of the NR1–NR2A ligand-binding

¹Department of Biochemistry and Molecular Biophysics and ²Howard Hughes Medical Institute, Columbia University, 650 West 168th Street, New York, New York 10032, USA.

†Present addresses: Vollum Institute, Oregon Health and Science University, Portland, Oregon 97239, USA (H.F., S.K.S., E.G.); The Structural Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, 10021, USA (R.M.).

core identifies a site in the heterodimer interface that has a key role in modulating the rate of receptor deactivation.

Structures of the ligand-binding cores

NMDA receptors are modular proteins, and the ligand-binding S1S2 domain can be prepared as a water-soluble protein that binds agonists and antagonists with affinities similar to those of the full-length assembled receptors²⁸ (Fig. 1). In the AMPA subtype of glutamate receptor, the GluR2 S1S2 domain participates in key subunit–subunit contacts in the intact receptor. To understand the molecular basis for intersubunit interactions in NMDA receptors, we solved structures of the NR2A S1S2 domain bound to glutamate (Supplementary Table S1 and Fig. S1) and the NR1–NR2A S1S2 complex bound to glycine and glutamate (Fig. 2 and Supplementary Table S1). Together with previously determined structures of NR1 S1S2 (refs 28, 29), we now have atomic-scale views of the glycine- and glutamate-binding sites and, most importantly, we have the structure of the NR1–NR2A S1S2 heterodimer.

The NR1–NR2A S1S2 heterodimer crystallizes as a single complex in the asymmetric unit with the two subunits related by a pseudo two-fold axis located at the centre of the interface (Fig. 2). The heterodimer buries 2,640 Å² of solvent-accessible surface area, and the sites of subunit–subunit contact can be divided into three subsites: sites I and III are related by the pseudo two-fold axis, and site II is on the pseudo two-fold axis (Fig. 2). Site II includes Y535 and P532 of NR1, P527 of NR2A, a salt bridge between R755 of NR1 and E530 of NR2A, and a hydrogen bond between K531 of NR1 and the backbone carbonyl oxygen of F524 of NR2A (Fig. 2d). Sites I and III comprise primarily hydrophobic residues on helices D and J of

domain 1 and include I519, A524 and L777 of NR1 and I514, V526, L777 and L780 of NR2A (Fig. 2c, e). Sites I and III also include several polar contacts, not only between domain 1 of each subunit but also between domain 1 (helix J) and domain 2 (helix F). Because agonist binding results in lobe closure and movement of domain 2, the intersubunit contacts between domain 1 and domain 2, which are not seen in the GluR2 S1S2 dimer^{23,24,30}, provide a mechanism by which the binding of agonists can be coupled to the interactions between subunits, that is, to the dimer interface²¹.

The subunits in the NR1–NR2A S1S2 heterodimer are arranged in a ‘back-to-back’ fashion that is similar to the arrangement of subunits in the non-desensitized state of the GluR2 S1S2 homodimer^{23,24}. Indeed, domain 1 of each subunit of the heterodimer superimposes with a root-mean-square (r.m.s.) deviation of 0.9 Å onto the 256 corresponding Cα atoms of the GluR2 S1S2–glutamate structure²³ (Fig. 2f, g). By contrast, domain 2 of the NR1–NR2A S1S2 and the GluR2 S1S2 structures superimpose poorly owing to differences in domain closure and the structural organization of domain 2 (Fig. 2f). Nevertheless, the similar subunit arrangement in the NR1–NR2A S1S2 heterodimer and the GluR2 S1S2 homodimer has three important implications.

First, it shows that the basic arrangement and interactions of subunits in the ligand-binding cores is conserved between NMDA receptors and non-NMDA receptors. Second, because the S1S2 dimer interface is a crucial site for receptor modulation in GluR2 (ref. 24), it is plausible that the heterodimer interface in NR1–NR2A is a locus for NMDA receptor modulation. Third, the conserved architecture suggests that gating of the ion channel in NMDA receptors occurs by a mechanism similar to that in non-NMDA receptors; that is, closure of each ligand-binding core ‘clam shell’ results in separation of the linker regions proximal to the ion channel domain. Notably, the separation of the equivalent ‘linker’ positions in the NR1–NR2A–glycine–glutamate complex is ~32 Å, whereas this distance in the GluR2–glutamate complex is ~38 Å. The molecular basis for this difference stems from disparities in the conformations of domain 2 and in the extent of domain closure between the NR1–NR2A and GluR2 S1S2 structures.

Glutamate-binding site of NR2A

The crystal structure of the NR2A ligand-binding core reveals recognition elements for glutamate and suggests a mechanism by which the NR2 subunits bind NMDA (Supplementary Fig. S1). On comparing the agonist-binding site of NR2A with the corresponding site of GluR2 (refs 23, 30, 31), GluR5 (ref. 32) and GluR6 (refs 32, 33), we find that the crucial difference is a negatively charged residue that participates in binding the positively charged amino group of the agonist. In NR2A this residue is D731, whereas in non-NMDA receptors the equivalent residue is a glutamate. Because the aspartate present in NR2A is one methylene group shorter than the glutamate present in non-NMDA receptors, there is no salt bridge between D731 and the amino group of the agonist glutamate as there is in non-NMDA receptors (Supplementary Fig. S1c). In NR2A, the amino group of the agonist forms water-mediated hydrogen bonds to residues E413 and Y761 (Supplementary Fig. S1b). Because the NR2 subunits have the shorter aspartate residue, NMDA can be modelled into the agonist-binding pocket by displacing the water molecule W2 (Supplementary Fig. S1d). If a similar exercise is carried out in the context of GluR2, steric clash occurs between the *N*-methyl group and the glutamate residue of the receptor.

A second feature of the binding site in NR2A that is distinct from non-NMDA receptors is a van der Waals contact between the γ -carboxylate group of glutamate and Y730, a residue that is conserved among NR2 subunits (Supplementary Fig. S1b). This contact, as well as an interdomain hydrogen bond between Y730 and E413, may partially account for the high-affinity binding of glutamate to NMDA receptors. Accordingly, mutation of the equivalent tyrosine in NR2B to alanine increases by 450-fold

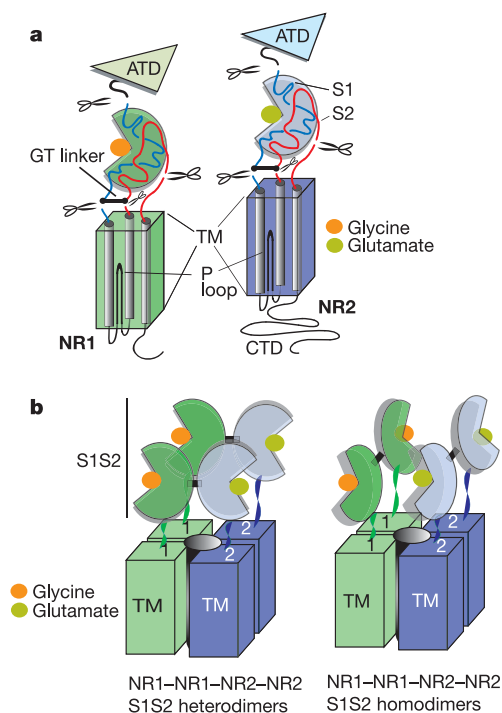


Figure 1 | Oligomeric arrangement in NMDA receptors. **a**, Domain organization of NMDA receptor subunits. Both NR1 and NR2 consist of N- (ATD) and C- (CTD) terminal domains, a transmembrane domain (TM) and a S1S2 ligand-binding core. The ligand-binding core can be isolated by tethering S1 and S2 with a Gly–Thr (GT) dipeptide linker. **b**, NMDA receptors form tetrameric channels comprising two copies each of NR1 and NR2. Shown here are the two possible modes of dimerization at the S1S2 ligand-binding cores, assuming that the subunits are organized in a NR1–NR1–NR2–NR2 arrangement²⁵. Thick black lines between the S1S2 domains indicate the formation of a particular homo- or heterodimer interface. ATD and CTD have been omitted for clarity.

the effector concentration for half-maximum response (EC_{50}) for glutamate³⁴. Despite differences in the pharmacology of glutamate binding and in the architectures of the glutamate-binding pockets, both NMDA receptors and non-NMDA receptors bind glutamate in the 'folded' rather than the 'extended' conformation, as suggested previously³⁵.

Physiological arrangement of NR1 and NR2A subunits

To validate the relevance of the NR1–NR2A subunit arrangement observed in the crystal to the intact receptor, we carried out biochemical, electrophysiological and sedimentation studies. We first engineered double cysteine mutants into both NR1 and NR2A by using the heterodimer crystal structure as a guide. Cysteines introduced at N521 and L777 of NR1 and E516 and L780 of

NR2A, sites related by the pseudo two-fold axis, are predicted to form unstrained disulphide bonds across the heterodimer interface (Fig. 3a). We reasoned that if the heterodimer interface is present in the intact receptor, then coexpression of the double cysteine mutants should result in crosslinked NR1 and NR2A subunits. By contrast, coexpression of the wild-type subunits, or subunits with only one set of double cysteine mutations, should not give rise to a disulphide-linked dimer.

Coexpression of wild-type subunits (WT–WT), NR1 double cysteine mutant and NR2A wild-type subunits (M–WT), or NR1 wild-type and NR2A double cysteine mutant subunits (WT–M) gave rise to bands that migrated as monomers under both reducing and non-reducing conditions (Fig. 3b). However, coexpression of the NR1 double cysteine mutant and NR2A double cysteine mutant

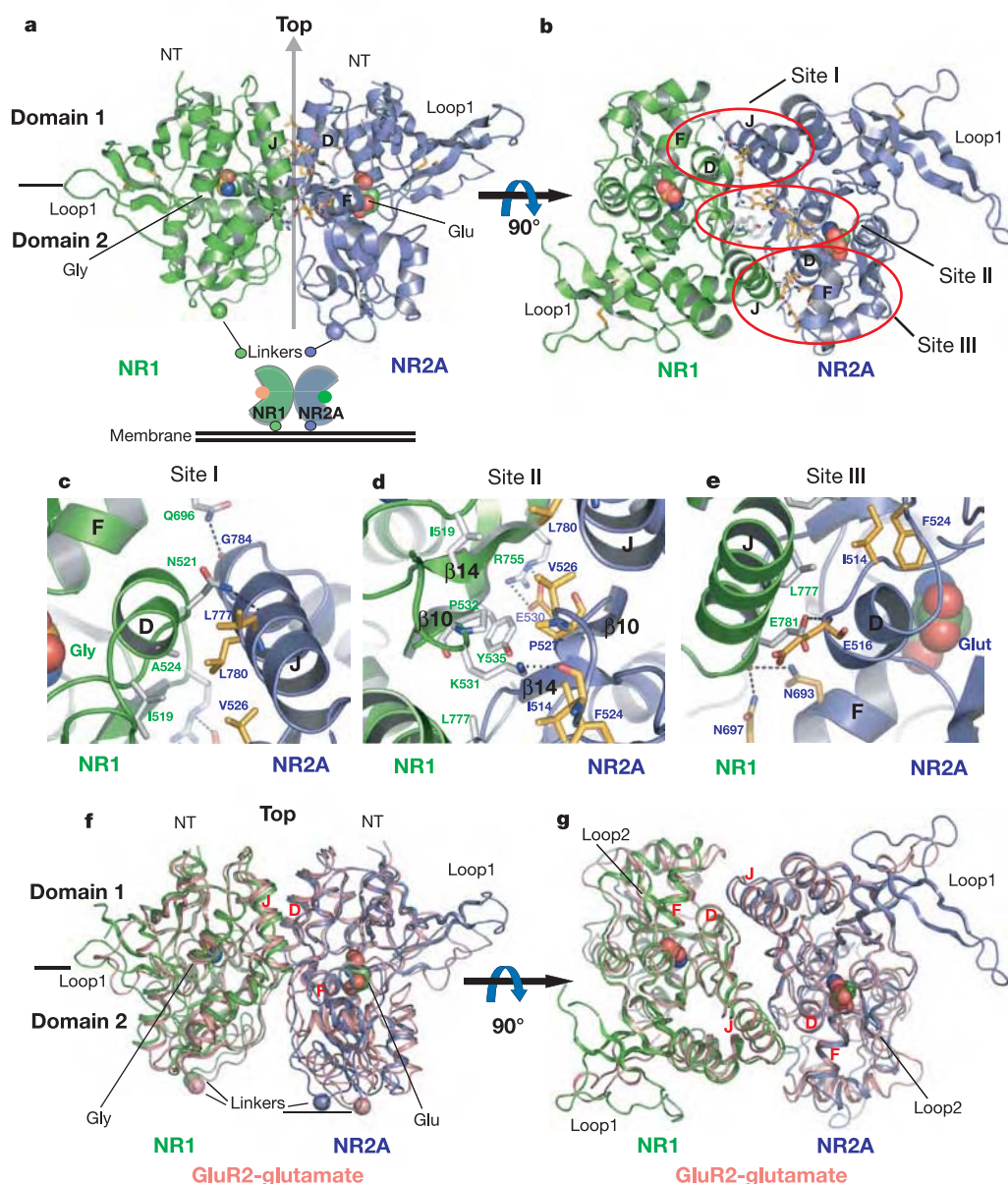


Figure 2 | Structure of NR1–NR2A S1S2. **a**, Side view of the NR1–NR2A S1S2 heterodimer in complex with glycine and glutamate. NR1 and NR2A are coloured green and blue, respectively. Glycine, glutamate and the C α atom of the glycine residue in the Gly–Thr dipeptide linker are shown as spheres. The arrow indicates the pseudo two-fold axis between the protomers. **b**, View of the structure from the 'top'. The interface between NR1 and NR2A is sliced into three sections denoted sites I–III. **c–e**, Magnified view of the interactions at sites I, II and III. Dashed lines

indicate hydrogen bonds or salt bridges. The interacting residues from NR1 and NR2A are coloured white and orange, respectively. **f**, **g**, Structural comparison between the NR1–NR2A (green–blue) S1S2 heterodimer and the glutamate-bound GluR2 S1S2 (pink) homodimer (PDB code 1FTJ). Superimposed structures are viewed from the side and 'top' of the molecules in **f** and **g**, respectively. Superposition was carried out on 256 residues from domain I with the program LSQKAB⁵⁰. The C α atoms of the glycine residues in the Gly–Thr dipeptide linkers are shown as spheres.

subunits (M–M) gave rise to a prominent band at a relative molecular mass of $\sim 300,000$ ($M_r \approx 300\text{K}$) that was recognized by antibodies against both NR1 and NR2A and was observed only under non-reducing conditions. We suggest that the $\sim 300\text{K}$ band corresponds to a spontaneously disulphide-linked NR1–NR2A heterodimer, in which the disulphide bridges form across the heterodimer interfaces of the S1S2 ligand-binding cores.

To determine whether the engineered cysteine residues affected the gating behaviour of the receptor, we measured ion channel activity for all wild-type and mutant combinations by two-electrode voltage clamp (TEVC). Application of glycine and glutamate elicited currents for all subunit combinations, indicating that the cysteine mutations did not disrupt receptor function. Application of 2 mM DTT for 1 min slightly potentiated the current induced by glutamate and glycine in the WT–WT channel, the effect of which was most probably due to reduction of the NR1 C744–C798 disulphide bond³⁶. Similar effects were seen in the WT–M and M–WT channels. By contrast, DTT significantly potentiated the current in the M–M channel (Fig. 3c). The extent of this potentiation was greater than the sum of the effects in the WT–M and M–WT channels, indicating that it was specific to the M–M combination of receptor subunits. Because the M–M combination gave rise to both a heterodimer band under non-reducing conditions in western blotting and the greatest DTT-mediated potentiation in the TEVC experiments, we propose that the arrangement of subunits in the NR1–NR2A S1S2 crystal structure is similar to that in the intact NR1–NR2A NMDA receptor.

To probe further the NR1–NR2A heterodimer interface, we measured the homo- and heterodimerization propensities of NR1 and NR2A S1S2 constructs by size-exclusion chromatography coupled to light scattering, refractive index and ultraviolet measurements (SEC-LS/RI/UV)³⁷, combined with sedimentation equilibrium and velocity experiments (Fig. 4 and Supplementary

Fig. S2). We found that wild-type NR1, NR2A or an equimolar mixture of NR1–NR2A S1S2 did not oligomerize to an appreciable extent, even at concentrations of up to 3 mg ml^{-1} (Fig. 4d and Supplementary Fig. S2 and Table S2), similar to the behaviour of the wild-type GluR2 S1S2 ligand-binding core²⁴. In GluR2, however, the mutation of L483 to tyrosine enhances S1S2 dimerization by 10^5 -fold and greatly slows receptor desensitization³⁸. Because the NR1–NR2A and GluR2 dimer interfaces are similar, we reasoned that by introducing tyrosine residues into the NR1 and NR2A subunits at positions equivalent to L483 in GluR2, we might be able to study the oligomerization behaviour of the NMDA receptor ligand-binding cores at protein concentrations amenable to light scattering and sedimentation analyses.

As predicted, the association pattern of the S1S2 proteins were altered by the introduction of these tyrosine residues (N521Y in NR1 and E516Y in NR2A), while ion channel activity was not significantly perturbed, as measured by TEVC (Supplementary Fig. S3 and Table S3). Specifically, SEC-LS/RI/UV and sedimentation equilibrium experiments showed that NR1 S1S2 N521Y dimerized with a dissociation constant of approximately 0.7 mg ml^{-1} (Fig. 4e, f and Supplementary Table S2), whereas NR2A S1S2 E516Y remained exclusively monomeric at concentrations up to 1.2 mg ml^{-1} (Fig. 4e, g, and Supplementary Table S2). However, when NR2A S1S2 E516Y was mixed with an equimolar ratio of NR1 S1S2 N521Y at a total protein concentration of 0.8 mg ml^{-1} and examined by SEC-LS/RI/UV, only a dimeric species was observed (Fig. 4e). The dramatic change in the association behaviour of NR2A S1S2 E516Y in the presence of NR1 S1S2 N521Y indicates that the NR2A ligand-binding core preferentially forms a heterodimer with the NR1 ligand-binding core. These results, together with the disulphide crosslinking and electrophysiological data, reinforce the conclusion that the NR1–NR2A arrangement seen in the crystal structure is present in the full-length receptor.

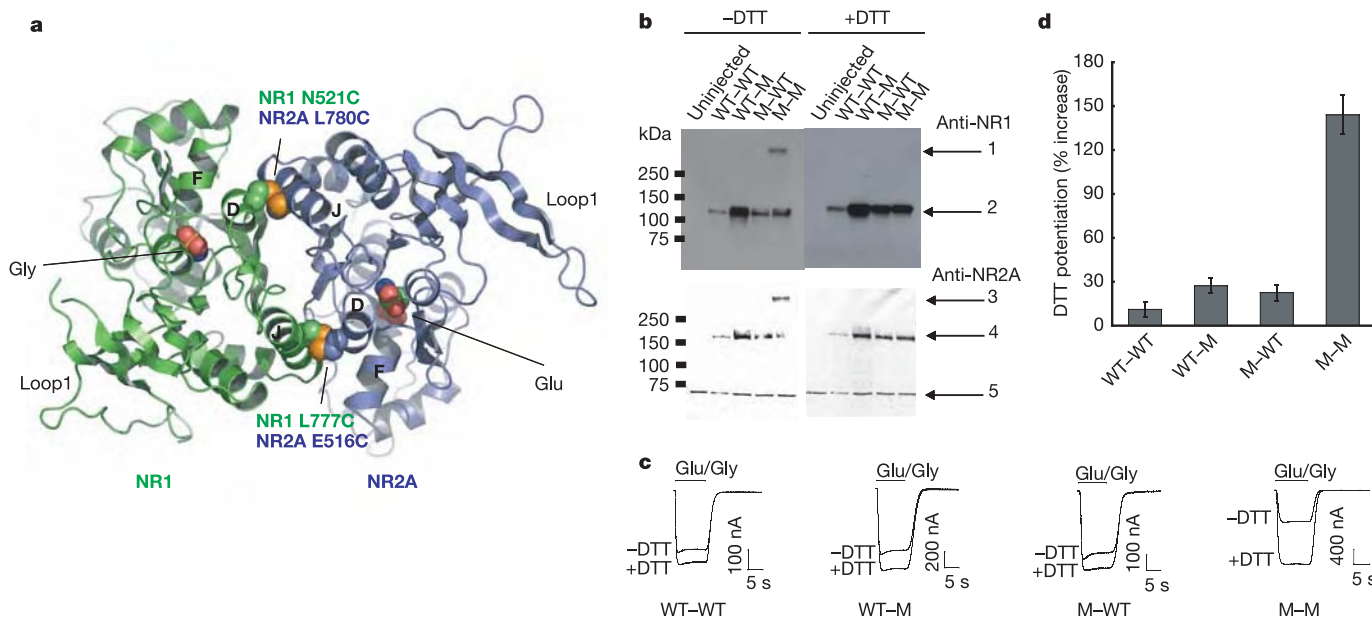


Figure 3 | Engineering disulphide bonds at the NR1–NR2A heterodimer interface. **a**, Location of the cysteines (spheres) engineered on helices D and J at positions N521 and L777 in NR1 and E516 and L780 in NR2A.

b, Assessment of disulphide bond formation by western blot. Membrane fractions of *Xenopus* oocytes expressing different combinations of wild-type (WT) and mutant (M) NR1–NR2A channels (WT–WT, M–WT, WT–M and M–M) were probed by antibodies against NR1 or NR2A in the presence or absence of DTT. Higher M_r species (arrows 1 and 3), representing the NR1–NR2A heterodimer, are seen only in the M–M lane ((+DTT)), and monomers (arrows 2 and 4) are seen in all of the other lanes ($\pm$ DTT). Bands

indicated by arrow 5 are nonspecific background. Enhanced chemiluminescence and alkaline phosphatase methods were used to detect bands on the anti-NR1 and anti-NR2A blots, respectively. **c**, **d**, Current recordings of the mutant channels in the presence of $300\text{ }\mu\text{M}$ glycine and $300\text{ }\mu\text{M}$ glutamate and in the presence or absence of DTT (2 mM for 1 min) by TEVC at -60 mV . On application of DTT significant potentiation ($\sim 140\%$ increase) was observed in the M–M channel ($n = 9$, $P < 0.001$), whereas only a small effect was observed in the WT–WT ($n = 4$), WT–M ($n = 6$) and M–WT ($n = 5$) channels. No significant run down was detected in the experiment. Error bars in **d** represent the s.d.

Site II in heterodimer interface modulates deactivation

The structural parallels between the NR1–NR2A S1S2 heterodimer and the GluR2 S1S2 homodimer suggest that specific regions in the intersubunit interface may have similar functions. Analysis of the GluR2 dimer interface shows that the regions corresponding to sites I and III (Fig. 2), namely the contacts between helix D on one subunit and helix J on the other, are involved in modulating receptor desensitization²⁴. By contrast, the region equivalent to site II (Figs 2 and 5) in AMPA receptors defines the binding site of small molecules, such as aniracetam, that modulate receptor deactivation³⁹ by stabilizing the ligand-binding domain in a closed-cleft, activated conformation. Interdomain hydrogen bonds stabilize a closed-cleft conformation^{23,32}, but binding of a modulator also does so by locking the ‘clam shell’ hinge³⁹. Comparison of site II in the NR1–NR2A S1S2 interface with the equivalent region in GluR2 shows that key conserved proline residues, P532 (NR1) and P527 (NR2A), are located in positions similar to those of the prolines in GluR2 (P494). Furthermore, the superposition shows that the aromatic ring of Y535 in NR1 occupies a position in the heterodimer equivalent to that of aniracetam in the GluR2 homodimer (Fig. 5),

namely at the hinge of the ligand-binding core clam shell. Specifically, the aromatic ring and hydroxyl group of NR1 Y535 overlap with one edge of the benzoyl ring and the adjacent edge of the pyrrolidinone ring of aniracetam, and the hydroxyl group of Y535 superimposes with the benzoyl carbonyl oxygen. Consequently, the tyrosine ring of NR1 Y535 and aniracetam in the GluR2 complex are involved in analogous interactions with their respective receptors. Because Y535 occupies a similar position in a NR1/NR2A heterodimer as aniracetam occupies in a GluR2 homodimer, we hypothesize that residue 535 modulates deactivation in NMDA receptors.

To test the role of NR1 Y535 in modulating NMDA receptor deactivation, we individually mutated this tyrosine to alanine, serine, leucine, phenylalanine and tryptophan, expressed the resulting receptor variants in HEK293 cells, and determined the rates of glycine- and glutamate-dependent deactivation by patch-clamp, rapid-solution-exchange measurements. For the wild type and all of the mutants, the rates of glycine and glutamate deactivation after brief applications of agonist were fit by a double exponential expressed with a fast (τ_f) and a slow (τ_s) component (Fig. 6 and Supplementary Table S4).

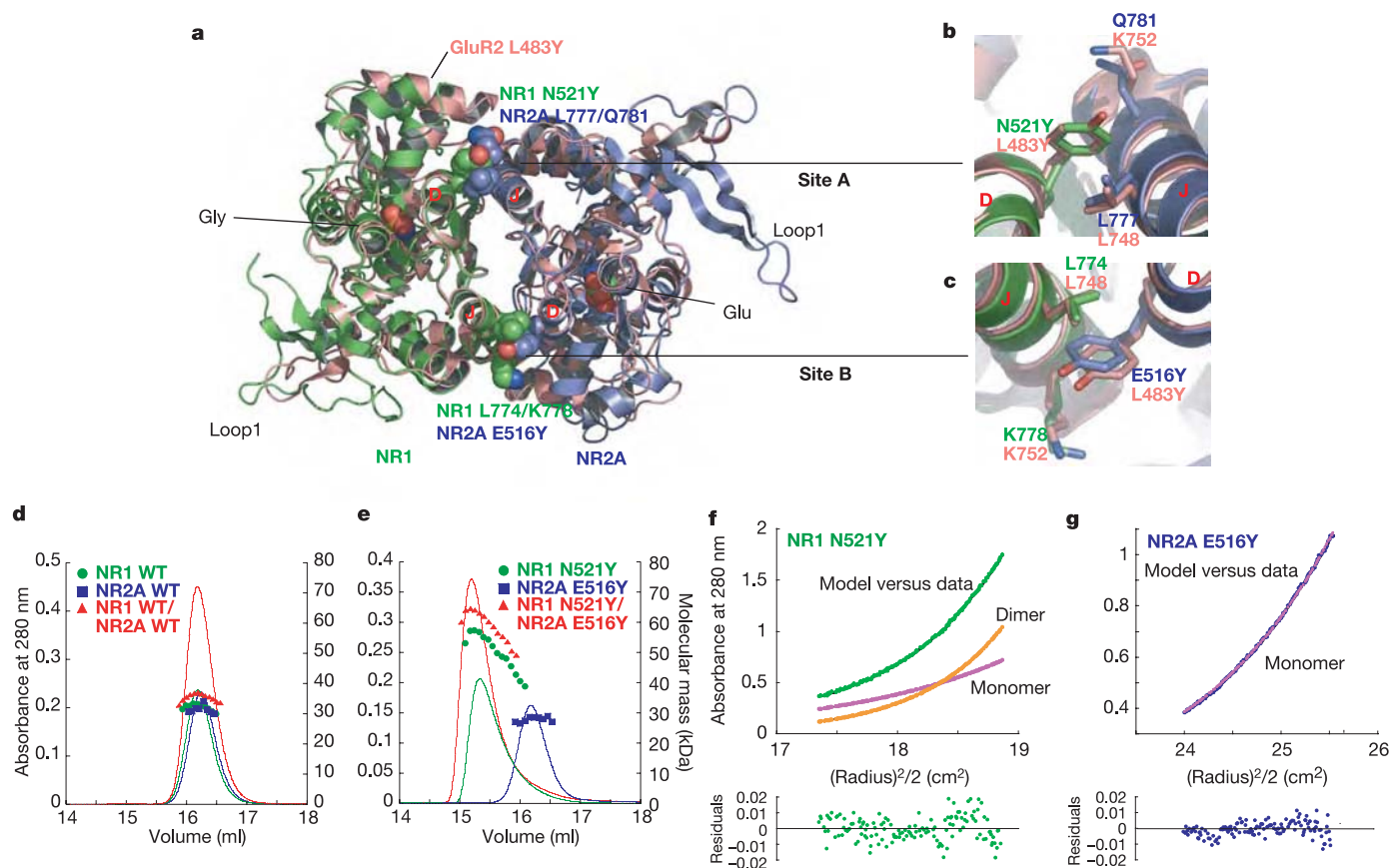


Figure 4 | Heterodimerization is favoured in NR1 S1S2 N521Y and NR2A S1S2 E516Y. **a**, Superposition of the GluR2 S1S2 L483Y (pink) homodimer onto the NR1–NR2A S1S2 (green–blue) heterodimer. The mutations N521Y in NR1 and E516Y in NR2A were designed using the GluR2 S1S2 L483Y structure (PDB code 1LB8) as a guide. **b, c**, Putative contacts at sites A and B in the NR1–NR2A mutant involve hydrophobic and cation– π interactions similar to those observed in GluR2 L483Y. **d, e**, Mass analysis by SEC-LS/RI/UV. The calculated M_r values are plotted for NR1 alone, NR2A alone and an equimolar mixture of NR1 and NR2A for both wild-type (**d**) and mutant (**e**) receptors. Solid lines represent the SEC profiles observed by absorbance at 280 nm. Dimeric and monomeric species form with NR1 S1S2 N521Y alone or with an equimolar mixture of NR1 S1S2 N521Y and NR2A S1S2 E516Y, whereas only monomeric species form with NR2A S1S2 E516Y or wild-type receptors. The concentration of all samples

was 0.8 mg ml^{-1} . **f, g**, Sedimentation equilibrium analysis of NR1 S1S2 N521Y (**f**) and NR2A S1S2 E516Y (**g**). Scans from three concentrations (0.2 , 0.4 and 0.8 mg ml^{-1}) and three rotor speeds ($13,000$, $18,000$ and $25,000 \text{ r.p.m.}$) were globally fit to either a monomer–dimer (NR1) or a single-species monomer (NR2A) model. Floating the reduced molecular weights yielded an M_r of 32.9 K and 31.6 K for the NR1 and NR2A monomers, respectively, which are within 1.5% of the values expected from their amino acid sequences disregarding disulphide bonds. Black lines indicate the model used to fit the data; green (NR1 N521Y) and blue (NR2A E516Y) circles indicate actual measurements; purple and orange lines represent the respective proportions of monomer and dimer calculated from the models. The graphs show the species distribution for the 0.8 mg ml^{-1} samples at $13,000 \text{ r.p.m.}$ Residuals in absorbance units are shown below each graph.

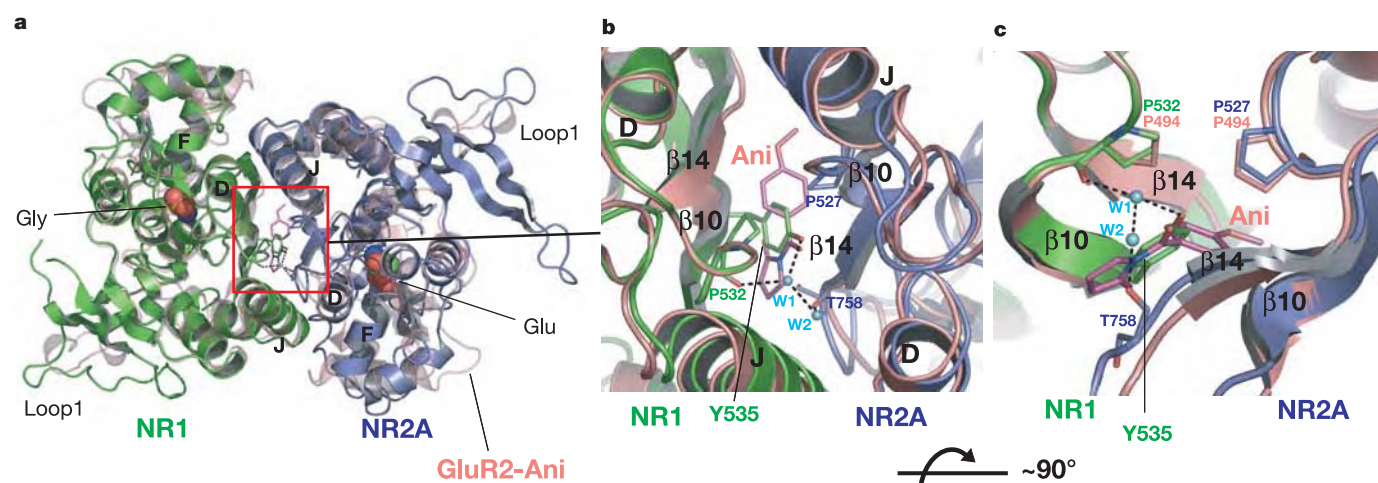


Figure 5 | Superposition of NR1-NR2A S1S2 and the GluR2 S1S2-aniracetam complex. **a**, Overlay of the GluR2 S1S2 dimer bound to glutamate and aniracetam (Ani, pink) onto the NR1-NR2A S1S2 dimer (green and blue) viewed from the same angle as in Fig. 1b. **b**, Magnification of the NR1 Y535 site and the aniracetam-binding site viewed from the same

angle as in **a**. Two water molecules, W1 and W2 (cyan spheres), participate in stabilizing the NR1-NR2A interaction. **c**, Side view of the NR1 Y535 site. Note that the position of the aniracetam molecule (pink) overlaps with that of the aromatic side chain of NR1 Y535.

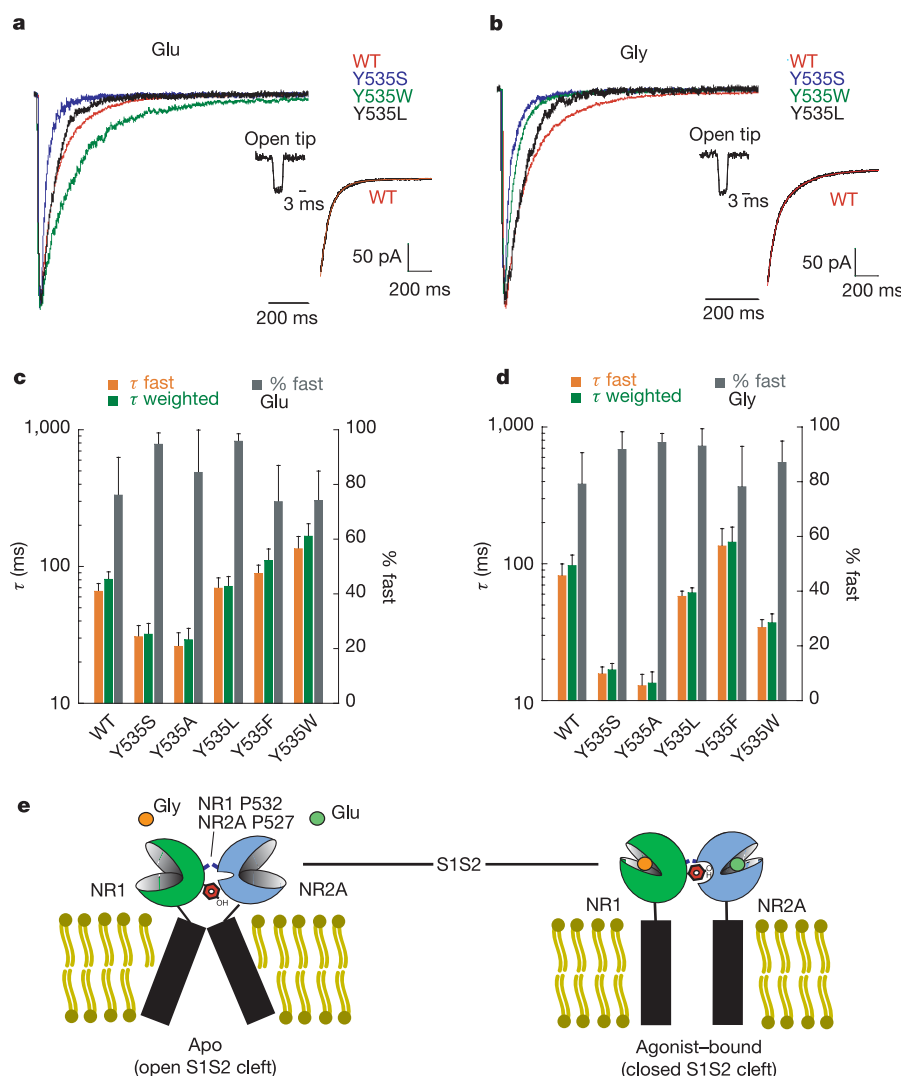


Figure 6 | Residue NR1 Y535 modulates the rate of NMDA receptor deactivation.

a, b, Normalized traces of glutamate- (**a**) or glycine- (**b**) induced currents for NR1 Y535S, Y535W, Y535L and wild-type (WT) Y535 combined with wild-type NR2A. Insets show the fit of the wild-type current decays to a double exponential equation after a 3-ms application of ligand (traces are shown in black and equation fits in red; see Methods). The typical open-tip response is roughly 500 μ s for a rise from 10 to 90%. **c, d**, Time constant values (τ) and percentages of the fast components of deactivation for glutamate- and glycine-induced currents calculated from the double exponential fit. Bars show mean $\pm$ s.d. from ten different patches. Note that the y axis for τ is on a log scale. **e**, Proposed mechanism by which Y535 of the NR1 subunit (red) slows the deactivation of NMDA receptors. The aromatic side chain of Y535 binds to a primarily hydrophobic pocket at the hinge region of the NR2A subunit, stabilizing the activated, glutamate-bound conformation. A single heterodimer with S1S2 and transmembrane domains is shown for clarity.

Mutation of NR1 Y535 to alanine or serine results in an increase in the rates of deactivation for glycine and glutamate by as much as 5–7-fold (Fig. 6c, d), owing more to an increase in the fast deactivation component (τ_f) and its relative weight (% fast) and less to a change in the rate of the slow component (τ_s). In AMPA receptors, the rate of deactivation is about two orders of magnitude greater than that in the NR1–NR2A NMDA receptor and the residue equivalent to NR1 Y535 is a serine, which suggests that the shape and volume of the residues at this position modulate the rate of receptor deactivation. The increase in the rate of glycine deactivation for the NR1 Y535S mutant is due, at least in part, to the fact that this mutant has a lower affinity or a greater off-rate for glycine than has the wild-type variant; that is, the glycine inhibition constant (K_i) for the soluble NR1 S1S2 Y535S mutant is roughly 2.5-fold higher than that for the wild-type construct (Supplementary Fig. S4). Because most NR1–NR2A receptors are present in a non-desensitized state after a brief (1–5 ms) application of ligand, dissociation (rather than entry and residence in the desensitized state) is most probably the main determinant of the deactivation rate in this case^{12,40}.

To explore the dependence of glycine and glutamate deactivation rates on the nature of the side chain at residue 535 of NR1, we examined the Y535L and Y535F mutants. Substitution of the aromatic ring by an aliphatic side chain in the Y535L mutant yields only a modest increase in the rates of glycine and glutamate deactivation, primarily owing to an increase in the weight of the fast component with no significant difference in τ_f and τ_s (Fig. 6c, d). This result underscores the importance of a planar, aromatic ring at position 535. In the crystal structures, the planar ring of aniracetam in GluR2 and Y535 in NR1 interact with conserved prolines (P494 in GluR2, P532 in NR1 and P527 in NR2A) through van der Waals contacts (Fig. 5c), interactions that are apparently crucial for slowing deactivation. Such interactions are absent in the rapidly deactivating non-NMDA receptors, which have small and hydrophilic residues (S497 in GluR2 and T504 in GluR6) at the position equivalent to Y535 in NR1. To a large extent, an aromatic residue at position 535 is important for maintaining the slow component of receptor deactivation. Accordingly, the Y535F mutant shows slightly slower deactivation rates and has weights for the fast and slow components similar to those of the wild-type receptor. This mutant also suggests that the hydroxyl group of NR1 Y535 and the two water-mediated hydrogen bonds (Fig. 5b, c) are not crucial in modulating the rate of receptor deactivation.

The tryptophan mutant, Y535W, has an aromatic side chain similar in size to aniracetam and shows two opposing effects: on the one hand, it slows the rate of glutamate-dependent deactivation by ~2-fold; on the other, it increases the rate of the glycine-dependent deactivation by ~2.6-fold (Fig. 6). If a tryptophan is modelled at position 535, the indole ring makes additional hydrophobic contacts with NR1 P532 and NR2A P527, and these contacts may stabilize the NR2A clam shell in a closed-cleft, activated conformation more effectively than does tyrosine. Why then does the tryptophan mutation speed the rate of glycine-dependent deactivation? Inspection of the open-cleft, antagonist-bound structure of NR1 S1S2²⁸ superimposed onto the glycine-bound NR1 subunit in the NR1–NR2A complex shows that the tyrosine occupies a different position in the antagonist-bound form owing to movement of the C γ atom by ~1.6 Å and rotation of the C δ –C γ –C β –C α dihedral angle by ~25°. Therefore, either the tryptophan may stabilize the NR1 glycine-binding domain in an open cleft conformation, perhaps by interactions with I755 of NR2A, or it may simply destabilize the closed-cleft state.

Conclusion

This study defines the NR1–NR2 S1S2 heterodimer as the fundamental functional unit in NR1–NR2A NMDA receptors and, together with previous studies, shows that the agonist-binding domains of NMDA and non-NMDA receptors are organized as

dimeric units. This conservation of architecture in turn suggests the conservation of a gating mechanism in which the agonist-induced closure of each ligand-binding domain results in separation of the ion channel proximal portions of the receptors (the 'linkers') and opening of the ion permeation pathway. Lastly, in both NMDA and non-NMDA receptors, the dimer interface provides sites for the allosteric modulation of gating activity.

METHODS

Structure determination. All data sets were collected on beam line X4A at the National Synchrotron Light Source using a Quantum 4 charged-coupled device detector (ADSC). The data sets were indexed, scaled and merged with HKL2000⁴¹. The rat NR2A S1S2 structure was determined by combining phases from single-wavelength anomalous dispersion (SAD) data collected on a NaBr-soaked crystal at the bromine peak energy and from molecular replacement using the C α coordinates for domain 1 of the rat NR1 S1S2–glycine structure without loops 1 and 2 (T396–V409, C454–Q487, E497–Q536 and S756–S800) as a search probe. The bromide sites were found with SOLVE⁴² and molecular replacement was done with AmoRe⁴³. Phase combination and extension were accomplished with SIGMA⁴⁴ and DM⁴⁵, respectively. We determined the NR1–NR2A S1S2 structure by molecular replacement using a heterodimer search probe model built by superposing the structures of NR1 S1S2–glycine²⁸ and NR2A S1S2–glutamate onto the GluR2 S1S2 dimer structure in helix J.

Refinement was done with CNS⁴⁶. Iterative rounds of model building in the program O⁴⁷ were carried out by using $F_o - F_c$ omit maps, coupled with Powell minimization and individual B-factor refinement, until R_{free} converged. When R_{free} was below 30%, ligands (glycine for NR1 and glutamate for NR2A) and water molecules were added, and the model was further refined until R_{free} converged again.

Electrophysiology. Oocyte recordings of the rat NR1–NR2A NMDA receptors were done in a TEVC configuration by using agarose-tipped microelectrodes filled with 3 M KCl at a holding potential of –60 mV. The bath solution contained 5 mM HEPES, 100 mM NaCl, 2.8 mM KCl, 10 mM Tricine and 0.3 mM BaCl₂ (pH 7.3).

For the fast-perfusion experiments, HEK293 cells (TsA201 variant) attached to 12-mm poly-D-lysine coated glass coverslips were transfected with plasmids encoding NR1-1a (rat), NR2A (rat) and green fluorescent protein using Lipofectamine2000 (Invitrogen). Recordings were made 24–36 h after transfection at room temperature with an Axopatch 200B amplifier (Axon Instruments). Recording electrodes (3–6 M Ω) were filled with 110 mM potassium gluconate, 2.5 mM NaCl, 5 mM BAPTA and 10 mM HEPES-KOH (pH 7.4). The wash solution contained 150 mM NaCl, 2.5 mM KCl, 2 mM CaCl₂, 10 mM HEPES-NaOH (pH 7.4), 10 mM glucose, 10 mM Tricine and 0.1 mM glycine or glutamate. The ligand solution contained wash solution plus 1 mM of glutamate or glycine and 5 mM sucrose to increase visibility of the solution interface. The outside-out patch was placed in front of the double-barrel theta tubing mounted on a piezoelectric device (Burleigh). The mean 10–90% rise time for the open tip response was typically 500–1,000 μ s. Brief (3-ms) pulses of the ligand solutions were applied and 20–50 recordings were obtained at 5-s intervals at a holding potential of –70 mV. The data points were averaged and fitted with a double exponential equation, $I(t) = I_f \times \exp(-t/\tau_f) + I_s \times \exp(-t/\tau_s)$, where I_f and I_s are the amplitudes of the fast and slow decay components and τ_f and τ_s are their respective decay time constants used to fit the data. We calculated the weighted mean decay time constant by the formula: $\tau_w = I_f/(I_f + I_s) \times \tau_f + I_s/(I_f + I_s) \times \tau_s$.

Analytical ultracentrifugation. Sedimentation equilibrium experiments were done in a Beckman Coulter Optima XL-I analytical ultracentrifuge with absorbance optics and quartz windows. Before these runs, protein samples were purified by Superdex 200 gel filtration chromatography, concentrated and dialysed overnight against a buffer containing 20 mM sodium phosphate (pH 7.5), 150 mM NaCl, 1 mM EDTA and either 1 mM glycine (NR1 S1S2 variants) or 1 mM L-glutamate (NR2A S1S2 variants). The mixture was made by combining equimolar amounts of mutant NR1 and NR2A and then dialysing it overnight against the same buffer containing both 1 mM glycine and 1 mM L-glutamate.

Protein samples were loaded into six-sector, 12-mm charcoal-filled Epon centrepieces at 0.1, 0.2, 0.4, 0.8, 1.0 and 1.2 mg ml^{–1} and run at 13,000, 18,000, and 25,000 r.p.m. in an An50Ti rotor at 4 °C. We collected absorbance (280-nm) scans at 2-h intervals with a 0.001-cm spacing and ten replicates per point. Data were analysed by nonlinear regression in WinNONLIN⁴⁸. Solvent density and viscosity were calculated with Sednterp⁴⁹. Association constants for any monomer–dimer equilibria obtained from WinNONLIN were converted from absorbance ($K_{2,\text{abs}}$) to molar units ($K_{2,\text{M}}$) with the equation $K_{2,\text{M}} = K_{2,\text{abs}}(e/l)/2$, where l is the path

length of the cell (1.2 cm) and ϵ is the molar extinction coefficient at 280 nm (36,840 and 32,430 M⁻¹ cm⁻¹, respectively, for NR1 and NR2A, estimated from the amino acid sequences disregarding disulphide bonds).

Received 2 June; accepted 20 July 2005.

- Kandel, E. R., Schwartz, J. H. & Jessel, T. M. *Essentials of Neural Science and Behavior* 219–306 (Appleton & Lange, East Norwalk, Connecticut, 1995).
- Watkins, J. C. & Evans, R. H. Excitatory amino acid transmitters. *Annu. Rev. Pharmacol. Toxicol.* **21**, 165–204 (1981).
- Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. Cloning by functional expression of a member of the glutamate receptor family. *Nature* **342**, 643–648 (1989).
- Keinänen, K. *et al.* A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560 (1990).
- Boulter, J. *et al.* Molecular cloning and functional expression of glutamate receptor subunit genes. *Science* **249**, 1033–1037 (1990).
- Werner, P., Voigt, M., Keinänen, K., Wisden, W. & Seeburg, P. H. Cloning of a putative high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. *Nature* **351**, 742–744 (1991).
- Moriyoshi, K. *et al.* Molecular cloning and characterization of the rat NMDA receptor. *Nature* **354**, 31–37 (1991).
- Johnson, J. W. & Ascher, P. Glycine potentiates the NMDA response in cultured mouse brain neurons. *Nature* **325**, 529–531 (1987).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. Voltage-dependent block by Mg²⁺ of NMDA responses in spinal cord neurones. *Nature* **309**, 261–263 (1984).
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. *Nature* **321**, 519–522 (1986).
- Forsythe, I. D. & Westbrook, G. L. Slow excitatory postsynaptic currents mediated by N-methyl-D-aspartate receptors on cultured mouse central neurones. *J. Physiol.* **396**, 515–533 (1988).
- Vicini, S. *et al.* Functional and pharmacological differences between recombinant N-methyl-D-aspartate receptors. *J. Neurophysiol.* **79**, 555–566 (1998).
- Lester, R. A., Clements, J. D., Westbrook, G. L. & Jahr, C. E. Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature* **346**, 565–567 (1990).
- Cull-Candy, S., Brickley, S. & Farrant, M. NMDA receptor subunits: diversity, development and disease. *Curr. Opin. Neurobiol.* **11**, 327–335 (2001).
- Waxman, E. A. & Lynch, D. R. N-methyl-D-aspartate receptor subtypes: multiple roles in excitotoxicity and neurological disease. *Neuroscientist* **11**, 37–49 (2005).
- Benveniste, M. & Mayer, M. L. Kinetic analysis of antagonist action at N-methyl-D-aspartate receptors. Two binding sites each for glutamate and glycine. *Biophys. J.* **59**, 560–573 (1991).
- Clements, J. D. & Westbrook, G. L. Activation kinetics reveal the number of glutamate and glycine binding sites on the N-methyl-D-aspartate receptor. *Neuron* **7**, 605–613 (1991).
- Watanabe, M., Inoue, Y., Sakimura, K. & Mishina, M. Developmental changes in distribution of NMDA receptor channel subunit mRNAs. *Neuroreport* **3**, 1138–1140 (1992).
- Monyer, H., Burnashev, N., Laurie, D. J., Sakmann, B. & Seeburg, P. H. Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* **12**, 529–540 (1994).
- Laurie, D. J. & Seeburg, P. H. Ligand affinities at recombinant N-methyl-D-aspartate receptors depend on subunit composition. *Eur. J. Pharmacol.* **268**, 335–345 (1994).
- Erreger, K., Chen, P. E., Wyllie, D. J. & Traynelis, S. F. Glutamate receptor gating. *Crit. Rev. Neurobiol.* **16**, 187–224 (2004).
- Ayalon, G. & Stern-Bach, Y. Functional assembly of AMPA and kainate receptors is mediated by several discrete protein-protein interactions. *Neuron* **31**, 103–113 (2001).
- Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).
- Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
- Schorge, S. & Colquhoun, D. Studies of NMDA receptor function and stoichiometry with truncated and tandem subunits. *J. Neurosci.* **23**, 1151–1158 (2003).
- Meddows, E. *et al.* Identification of molecular determinants that are important in the assembly of N-methyl-D-aspartate receptors. *J. Biol. Chem.* **276**, 18795–18803 (2001).
- Regalado, M. P., Villarroel, A. & Lerma, J. Intersubunit cooperativity in the NMDA receptor. *Neuron* **32**, 1085–1096 (2001).
- Furukawa, H. & Gouaux, E. Mechanisms of activation, inhibition and specificity: crystal structures of the NMDA receptor NR1 ligand-binding core. *EMBO J.* **22**, 2873–2885 (2003).
- Inanobe, A., Furukawa, H. & Gouaux, E. Mechanism of partial agonist action at the NR1 subunit of NMDA receptors. *Neuron* **47**, 71–84 (2005).
- Jin, R., Horning, M., Mayer, M. L. & Gouaux, E. Mechanism of activation and selectivity in a ligand-gated ion channel: structural and functional studies of GluR2 and quisqualate. *Biochemistry* **41**, 15635–15643 (2002).
- Armstrong, N., Sun, Y., Chen, G. Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
- Mayer, M. L. Crystal Structures of the GluR5 and GluR6 ligand binding cores: molecular mechanisms underlying kainate receptor selectivity. *Neuron* **45**, 539–552 (2005).
- Nanao, M. H., Green, T., Stern-Bach, Y., Heinemann, S. F. & Choe, S. Structure of the kainate receptor subunit GluR6 agonist-binding domain complexed with domoic acid. *Proc. Natl Acad. Sci. U S A* **102**, 1708–1713 (2005).
- Laube, B., Schemm, R. & Betz, H. Molecular determinants of ligand discrimination in the glutamate-binding pocket of the NMDA receptor. *Neuropharmacology* **47**, 994–1007 (2004).
- Watkins, J. C., Krogsgaard-Larsen, P. & Honore, T. Structure-activity relationships in the development of excitatory amino acid receptor agonists and competitive antagonists. *Trends Pharmacol. Sci.* **11**, 25–33 (1990).
- Choi, Y., Chen, H. V. & Lipton, S. A. Three pairs of cysteine residues mediate both redox and Zn²⁺ modulation of the NMDA receptor. *J. Neurosci.* **21**, 392–400 (2001).
- Folta-Stogniew, E. & Williams, K. R. Determination of molecular masses of proteins in solution: implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J. Biomol. Technol.* **10**, 51–63 (1999).
- Stern-Bach, Y., Russo, S., Neuman, M. & Rosenmund, C. A point mutation in the glutamate binding site blocks desensitization of AMPA receptors. *Neuron* **21**, 907–918 (1998).
- Jin, R. *et al.* Mechanism of positive allosteric modulators acting on AMPA receptors. *J. Neurosci.* **25**, 9027–9036 (2005).
- Erreger, K., Dravid, S. M., Banke, T. G., Wyllie, D. J. & Traynelis, S. F. Subunit-specific gating controls rat NR1/NR2A and NR1/NR2B NMDA channel kinetics and synaptic signalling profiles. *J. Physiol.* **563**, 345–358 (2005).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Navaza, J. *Amore*: An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
- Cowan, K. D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 43–48 (1996).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**(5), 905–921 (1998).
- Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
- Yphantis, D., Johnson, M. L. & Lary, J. W. *WinNONLIN06* program. (National Analytical Ultracentrifugation Facility, Univ. Connecticut, Storrs, CT, 1997).
- Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
- Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Rowe for comments on the sedimentation equilibrium data analysis. M. Mayer is thanked for discussions, critical reading of this manuscript, and pSP NR1-1a and NR2A. J. Howe and S. Traynelis are thanked for comments on electrophysiological experiments and for the TsA201 cell-line and pCINEO NR1-1a and NR2A, respectively. We thank N. Armstrong, W. Zhang, and A. Robert for instructions on the patch-clamp and rapid solution exchange experiments; R. Abramowitz and X. Yang for assistance with the X-ray experiments; S. Siegelbaum for providing *Xenopus* oocytes; and A. Sobolevsky and O. Boudker for critically reading the manuscript. The NR1 and NR2A cDNAs used in the structural analysis were a gift from S. F. Heinemann. S.K.S. is supported by a NIH National Research Service Award postdoctoral fellowship. The Beckman XL-I analytical centrifuge was purchased with funds from the NIH. The work was supported by the NIH. E.G. is an investigator with the Howard Hughes Medical Institute.

Author Information The coordinates and structure factors for NR1-NR2A S1S2 and NR2A S1S2-glutamate have been deposited in the Protein Data Bank with accession codes 2A5T and 2A5S, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.G. (jeg52@columbia.edu).

Simulation of equatorial and high-latitude jets on Jupiter in a deep convection model

Moritz Heimpel¹, Jonathan Aurnou² & Johannes Wicht³

The bands of Jupiter represent a global system of powerful winds. Broad eastward equatorial jets are flanked by smaller-scale, higher-latitude jets flowing in alternating directions^{1,2}. Jupiter's large thermal emission suggests that the winds are powered from within^{3,4}, but the zonal flow depth is limited by increasing density and electrical conductivity in the molecular hydrogen–helium atmosphere towards the centre of the planet⁵. Two types of planetary flow models have been explored: shallow-layer models reproduce multiple high-latitude jets, but not the equatorial flow system^{6–8}, and deep convection models only reproduce an eastward equatorial jet with two flanking neighbours^{9–14}. Here we present a numerical model of three-dimensional rotating convection in a relatively thin spherical shell that generates both types of jets. The simulated flow is turbulent and quasi-two-dimensional and, as observed for the jovian jets, simulated jet widths follow Rhines' scaling theory^{2,12,13,15}. Our findings imply that Jupiter's latitudinal transition in jet width corresponds to a separation between the bottom-bounded flow structures in higher latitudes and the deep equatorial flows.

Turbulent energy typically passes from larger to smaller scales. However, this process can reverse under conditions favouring coherent flow. This is the case for rapid planetary rotation (geostrophy), where small-scale turbulent motions can feed large-scale zonal jets—a process that seems to be discernible from analysis of the jovian surface winds¹⁵. The effects of boundary curvature in a rotating system are quantified by a parameter β . The inverse cascade from small to large scales ceases at the Rhines length, which is inversely proportional to β and sets the characteristic width of zonal jets¹⁶. For a shallow planetary layer, with fluid motion constrained to the outer spherical surface, β depends upon the latitudinal variation of the local strength of planetary rotation^{12,13,16}. On the other hand, turbulent flow in a rapidly rotating spherical volume generates deep cylindrically symmetric flow structures aligned with the rotation axis. Deep zonal flow is governed by the topographic β -parameter, which depends on the gradient of the axial fluid column height, and has sign opposite to that of β for a shallow layer. This sign difference implies that a shallow layer produces retrograde (westward) equatorial flow as observed on Uranus and Neptune, whereas full-sphere dynamics results in a prograde (eastward) equatorial jet, as on Jupiter and Saturn^{6,12}.

The existence of an inner spherical boundary modifies deep zonal flow by separating it into three connected but dynamically distinct regions (north, equatorial and south), bounded by the tangent cylinder that circumscribes the inner boundary equator. The radius ratio $\chi = r_i/r_o$ sets the location of the tangent cylinder and is of fundamental importance to rapidly rotating convection. Figure 1 illustrates the structure of geostrophic zonal flow in a spherical shell. Inside the tangent cylinder, northern and southern flow structures are separated by the inner boundary. In contrast, flow structures

extend over both hemispheres in the equatorial region. The tangent cylinder therefore corresponds to a scaling discontinuity. Starting at the equator, the axial column height increases with latitude. Outside the tangent cylinder, scaling is identical to that of a full sphere, and the sign of β is consistent with a prograde equatorial jet. Crossing the latitude θ_{tc} , where the tangent cylinder intersects the outer surface, the column height is reduced by one-half, β changes sign, and fluid columns decrease in height toward the pole¹⁷. Because of this sign reversal, jet scaling inside the tangent cylinder is analogous to that of a shallow layer. (Indeed, the limiting case of a very thin three-dimensional spherical shell is equivalent to a two-dimensional layer.) Furthermore the jump in β decreases the Rhines length inside the tangent cylinder, favouring the formation of multiple jets there.

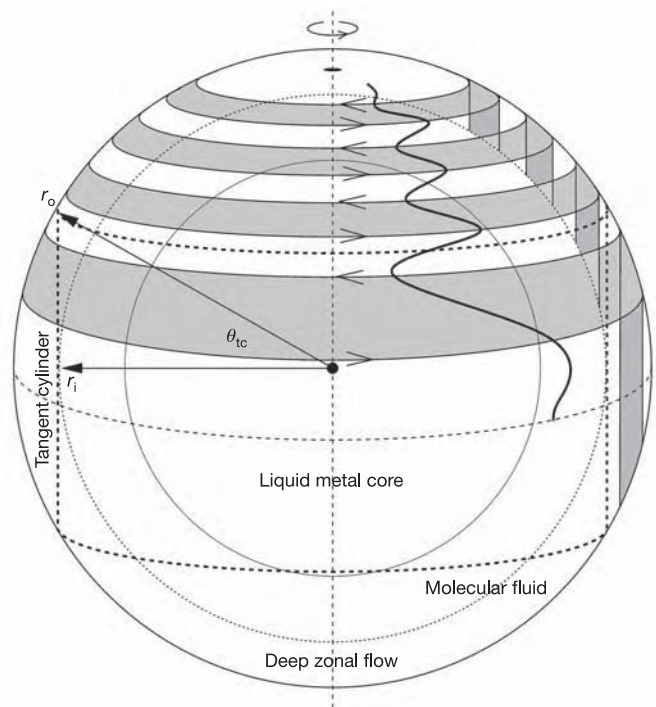


Figure 1 | Illustration of rapidly rotating turbulent convection in a spherical shell. Flow occurs between outer radius r_o and inner radius r_i . The latitudes $\theta_{tc} = \pm \cos^{-1}(r_i/r_o)$ mark the intersection of the tangent cylinder with the outer boundary. The shaded and white areas in the northern hemisphere correspond (on the outer surface) to the visible jovian belts and zones¹. The radius ratio shown here, and the size of the liquid metal core, are chosen for the purpose of illustration and do not represent estimates for Jupiter. The origin of Jupiter's visible colours is not addressed in this paper.

¹Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada. ²Department of Earth and Space Sciences, UCLA, Los Angeles, California 90095-1567, USA.

³Max Planck Institute for Solar System Research, 37191 Katlenburg-Lindau, Germany.

Previous analytical models of planetary zonal flow in a spherical shell have assumed a relatively deep bottom boundary, such that high-latitude zonal jets develop from flows outside the tangent cylinder¹⁸. More recent numerical models of deep and strongly turbulent three-dimensional quasigeostrophic convection^{10,11} have produced jets of realistic amplitudes. However, these moderately thick fluid shells (with radius ratios $\chi = r_i/r_o = 0.6\text{--}0.75$) produce only a pair of high-latitude jets in each hemisphere inside the tangent cylinder^{10,14}, which cannot account for the observed pattern of jovian jets. Laboratory experiments of rotating convection in deep spherical shells¹⁹ with $\chi = 0.35$ and 0.70 have obtained zonal flow patterns that are broadly comparable to the results of spherical numerical models^{9,10}. However, multiple jets have been produced by idealized numerical²⁰ and experimental²¹ models with a cylindrical geometry, a free top surface, and a sloping bottom surface. In those local models, as well as our present global model, the jets are produced by the topographic β -effect and follow Rhines scaling.

There are various possible reasons why previous spherical shell models have not produced multiple higher-latitude jets. Instead of analysing particular cases we list the following conditions that favour the development of multiple high-latitude jets. Turbulent flow (that is, high Reynolds number) constrained by rapid rotation (that is, low Rossby number) is necessary for the development of jets that follow Rhines scaling. A relatively thin fluid layer allows multiple jets to form at higher latitudes by decreasing the Rhines length and increasing the latitudinal range inside the tangent cylinder.

We use numerical modelling to study turbulent thermal convection in a rapidly rotating three-dimensional spherical shell, with simulation parameters chosen to reflect our present understanding of Jupiter's interior. The spherical shell geometry is defined by the radius ratio $\chi = r_i/r_o$. We have chosen $\chi = 0.9$, which represents a substantially thinner shell than in previous models of rotating convection^{10,14}. This value of the radius ratio corresponds to a depth in Jupiter of approximately 7,000 km, which is shallower than the phase boundary that separates the outer molecular fluid

from the liquid metal H–He core, estimated at $0.8\text{--}0.85 R_J$, where $R_J \approx 70,000$ km is the radius of Jupiter¹³. Measurements of increasing flow velocity with depth suggest that the surface winds are seated in the deep molecular H₂–He atmosphere²². However, increasing density and electrical conductivity with pressure result in inertial and Lorentz forces that are expected to damp the zonal flow between 0.85 and $0.95 R_J$ (refs 13 and 23). Thus we have chosen a simulation radius ratio that lies near the middle of current estimates. A description of Saturn's interior is similar to that of Jupiter except that lower saturnian gravity roughly doubles the estimated layer depth.

Selection of other simulation parameters (see Methods) is based on recent numerical and experimental scaling analyses for convection-driven zonal flows in thicker spherical shells^{11,24}. An essential ingredient in this numerical simulation is that convective turbulence is quasigeostrophic and close to the asymptotic state of rapid rotation in which viscosity and thermal diffusivity play a negligible role in the dynamics. Thus, large discrepancies between the simulation parameters and those estimated for Jupiter should not strongly affect the character of the solution. We do not model the jovian troposphere, nor the effects of latitudinally varying insolation^{7,8}. Furthermore, we model convection only within the region where large-scale zonal flows are predicted to occur and we neglect deeper regions where convection may be vigorous but zonal flows are expected to be weak. Although fluid compressibility effects are important to the dynamics of convection in the gas giants²⁵, the fluid in our numerical model is assumed to be incompressible except for thermal buoyancy effects (that is, the Boussinesq approximation). The omission of compressibility effects is possibly this model's greatest limitation. However, considering that we use a relatively thin convection layer, a Boussinesq treatment may be adequate to simulate the large-scale dynamics^{12,13}.

Figure 2 shows the results of our numerical simulation and the jovian zonal wind pattern. Figure 2a is a plot of Jupiter's averaged surface azimuthal (east–west) velocity profile relative to the deep-seated magnetic-field reference frame¹. Although we focus here on

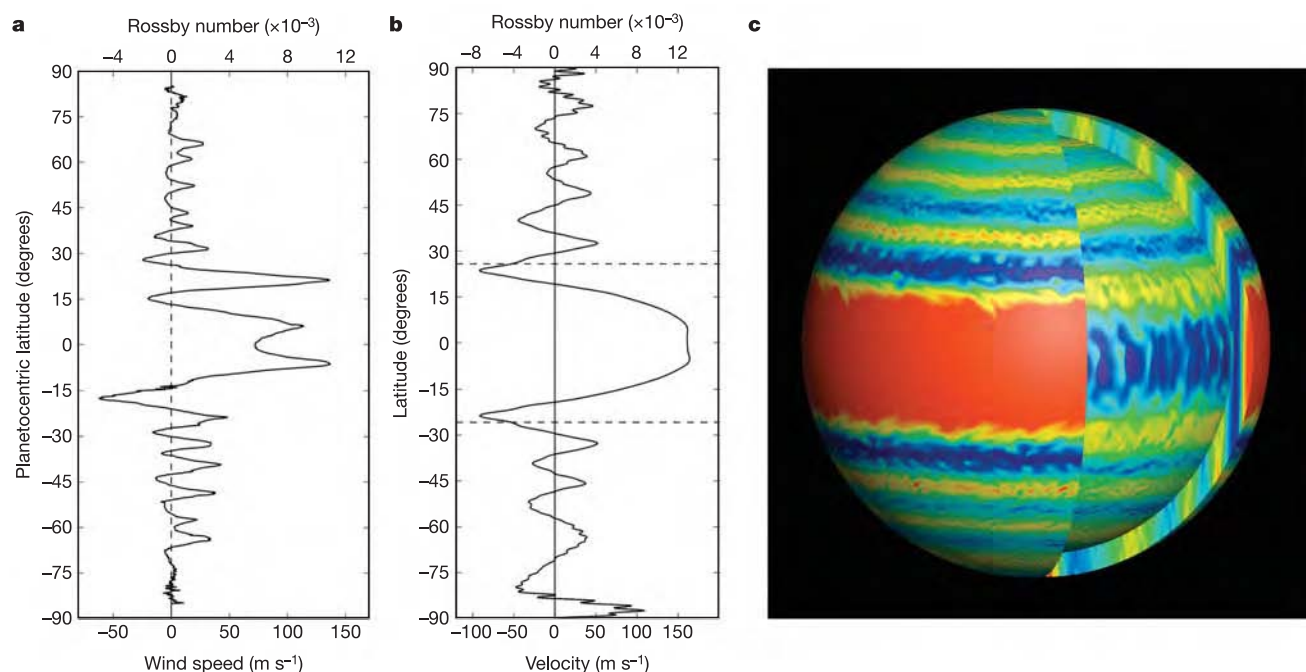


Figure 2 | Zonal flow for Jupiter and the numerical simulation. **a**, The zonal wind of Jupiter. The Cassini space mission data was kindly provided by A. Vasavada. Wind speed may be converted to Rossby number using $Ro = v/\Omega r_o$. **b**, Snapshot in time of the simulation's outer-surface azimuthally averaged azimuthal (east–west) velocity. Dashed lines indicate

the latitude where the tangent cylinder intersects the outer surface, for the simulation's radius ratio ($\chi = r_i/r_o = 0.9$). **c**, Snapshot of the simulation's azimuthal velocity field on the outer and inner spherical surface, and on a meridional slice. Red and blue colours represent prograde (eastward) and retrograde (westward) flow, respectively.

Jupiter, it is noted that Saturn has a comparable wind pattern, with a stronger and broader eastward equatorial jet and roughly half the number of high-latitude jets. Figure 2b and c shows aspects of the flow field resulting from the numerical model. In the simulation, a prominent equatorial jet is underlaid by flow structures that are aligned with the rotation axis and span the two hemispheres. At higher latitudes, multiple alternating jets form a series of banded flows that are underlaid by axially aligned flow structures bounded by the outer and inner spherical surfaces. The cylindrical symmetry of the simulation's global flow field shows that it is quasigeostrophic.

To test the applicability of the spherical shell model to Jupiter we compare the simulated and jovian jet scaling to that predicted by the theory of geostrophic turbulence. Analysis of Rhines scaling in spherical shell geometry (see Methods) results in predicted jet widths based on the mean zonal flow velocity within the three regions bounded by the tangent cylinder (north, equatorial and south). Predicted versus measured jet widths are plotted in Fig. 3. For both Jupiter and our numerical model, the jet widths are relatively narrow and constant at high latitudes and increase sharply toward the equator. The simulated equatorial jets are broader than predicted, which suggests that, in our simulation, Rhines scaling is not dominant outside the tangent cylinder. However, at higher latitudes the simulated jet widths clearly follow Rhines scaling, and the equatorward increase in jet widths occurs, as expected, at the latitude of the tangent cylinder. A similar scaling transition is also evident for the jovian jet widths. This scaling transition is predicted for turbulent convection in a rapidly rotating spherical shell, but not for a full sphere or a two-dimensional shallow layer.

Current understanding of Jupiter's zonal winds is based largely upon observations of the surface motions and heat flow. Additional constraints on the nature and extent of deep flows could come from accurate measurements of the planetary gravitational field. Deep zonal flows can carry significant angular momentum, enough to

generate measurable latitudinal gravity variations²⁶. Further evidence could involve higher-resolution observations of Jupiter's outward heat flux. Such evidence will be provided by future space missions, including the planned jovian polar orbiter Juno²⁷.

METHODS

Numerical modelling. The pseudo-spectral numerical code uses mixed implicit/explicit time stepping and has been benchmarked²⁸. Values of non-dimensional control parameters²⁸ for the simulation are as follows: $\chi = 0.9$, $E = 3 \times 10^{-6}$, $Ra = 1.67 \times 10^{10}$, $Pr = 0.1$. Output parameters are measured to be approximately $Re = 5 \times 10^4$ and $Ro = 0.012$. The radius ratio $\chi = r_i/r_o$, where r_i and r_o are the inner and outer boundary radii, defines the spherical shell geometry. The Ekman number E is the ratio of viscous and Coriolis forces. The Rayleigh number Ra gives the strength of buoyancy forces in the flow. The Prandtl number Pr is the ratio of viscous and thermal diffusion. The Reynolds number Re is the ratio of inertial and viscous forces. The Rossby number Ro is the ratio of inertial and Coriolis forces. The top and bottom velocity boundary conditions are mechanically impenetrable and stress-free. The thermal boundary conditions are fixed temperature. The initial conditions are zero motion relative to the rotating frame and a random thermal perturbation, from which convection develops. Solving the governing equations on an 8-fold azimuthally truncated sphere²⁹, we use 768 points in latitude, 192 points in longitude and 65 points in radius. Convergence of the numerical simulation required the use of hyperdiffusion, which was reduced as the calculation became steadier in time. The hyperdiffusion has the same functional form (but with 1/10 of the final strength) of previous dynamo models³⁰. The Ekman number was also lowered in stages. After E was reduced from 3×10^{-5} to 3×10^{-6} the model was run for over 1,600 planetary rotations and the convective motions approached a quasi-steady state. The total calculation time represents about 2,200 rotations and 0.14 viscous diffusion times.

Analysis of jet scaling. In meteorology and oceanography, the β -plane approximation is often used to represent flows in a very thin spherical shell by two-dimensional motions on a spherical surface. In this approximation, the Rhines length (which sets the latitudinal scale of zonal jets) is:

$$L_\beta = \left(\frac{V r_o}{\Omega \cos \theta} \right)^{1/2} \quad (1)$$

where θ is the latitude and V is a zonal flow velocity scale.

Convection in rapidly rotating deep spherical shells is largely invariant along the direction of the rotation axis, that is, in the z direction. The resulting flows are quasigeostrophic. In the limit of rapid rotation, flow becomes two-dimensional and is called geostrophic. The geostrophic Rhines length then refers to the scale in the cylindrical radius directions s :

$$L_g = \left(\frac{V h}{\Omega |\partial h / \partial s|} \right)^{1/2} \quad (2)$$

The Rhines length now depends on the variation of the spherical shell height $h(s)$, measured in the direction of the rotation axis. This height dependence is referred to as the topographic β -effect¹³. Figure 2 demonstrates that the flow of our numerical simulation is indeed dominated by strong quasi-geostrophic zonal flows: the flow has a high degree of cylindrical symmetry and depends mainly on cylindrical radius.

The scales L_β and L_g are similar in many ways: both are based on two-dimensional approximations for a three-dimensional system¹³. However, a profound difference between the β -plane and geostrophic approximations is the existence of a tangent cylinder, which represents a discontinuity in the Rhines scaling. The height $h(s)$ doubles discontinuously outward across the tangent cylinder, while $\partial h / \partial s$ changes sign. Furthermore, stretching of fluid columns inhibits flow across the tangent cylinder. This effectively isolates the flow fields, suggesting that zonal flow scaling can be analysed separately inside and outside the tangent cylinder.

Accordingly, we take the Rhines length to be constant within each of the three regions bounded by the tangent cylinder (north, equatorial and south). Using this assumption we have derived a precise mapping from the observed regional velocity V to the predicted regional Rhines wavelength, given here in radians of latitude:

$$\lambda_g = 2\pi \left(\frac{V}{r_o \Omega C(\chi)} \right)^{1/2} \quad (3)$$

where $C(\chi)$ is a geometrical parameter containing latitudinal mean changes in the spherical shell height for the distinct regions inside and outside the tangent cylinder (see Supplementary Information for details). The regional Rhines widths $\lambda/2$ are compared with the measured jet widths of Jupiter and the numerical simulation in Fig. 3.

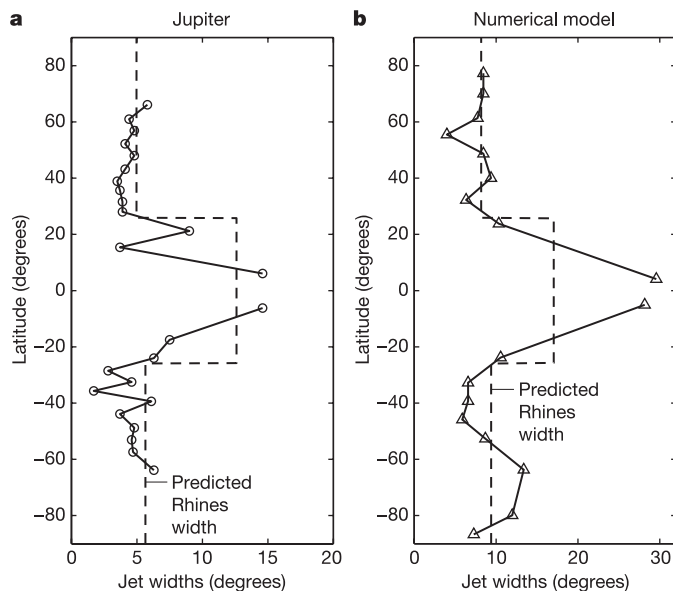


Figure 3 | Measured jet widths compared to jet widths predicted by Rhines scaling for Jupiter (a) and the numerical model (b). Circle and triangle symbols represent individual jet measurements (see Supplementary Information for jet-measurement method). Predicted widths (dashed black lines) are obtained from observed regional (north, equatorial, and south) zonal velocities (see equation (3) in the Methods section). The latitude of the discontinuity in predicted jet widths represents the intersection of the tangent cylinder with the outer surface (see Fig. 1) for the simulation's spherical shell radius ratio ($\chi = 0.9$). The same radius ratio was used to obtain predicted jets widths for Jupiter.

Received 18 May; accepted 1 September 2005.

1. Porco, C. C. *et al.* Cassini imaging of Jupiter's atmosphere, satellites and rings. *Science* **299**, 1541–1547 (2003).
2. Vasavada, A. R. & Showman, A. P. Jovian atmospheric dynamics: an update after Galileo and Cassini. *Rep. Prog. Phys.* **68**, 1935–1996 (2005).
3. Ingersoll, A. P. Pioneer 10 and 11 observations and the dynamics of Jupiter's atmosphere. *Icarus* **29**, 245–253 (1976).
4. Pirraglia, J. A. Meridional energy balance of Jupiter. *Icarus* **59**, 169–176 (1984).
5. Guillot, T., Stevenson, D. J., Hubbard, W. & Saumon, D. in *Jupiter, the Planet, Satellites and Magnetosphere* (eds Bagenal, F., Dowling, T. E. & McKinnon, W. B.) 35–57 (Cambridge Univ. Press, Cambridge, 2004).
6. Cho, J. Y.-K. & Polvani, L. M. The morphogenesis of bands and zonal winds in the atmospheres on the giant outer planets. *Science* **273**, 335–337 (1996).
7. Williams, G. P. Planetary circulations. 1. Barotropic representation of Jovian and terrestrial turbulence. *J. Atmos. Sci.* **35**, 1399–1426 (1978).
8. Williams, G. P. Jovian dynamics. Part III: Multiple, migrating, and equatorial jets. *J. Atmos. Sci.* **60**, 1270–1296 (2003).
9. Aurnou, J. M. & Olson, P. L. Strong zonal winds from thermal convection in a rotating spherical shell. *Geophys. Res. Lett.* **28**, 2557–2559 (2001).
10. Christensen, U. R. Zonal flow driven by deep convection on the major planets. *Geophys. Res. Lett.* **28**, 2553–2556 (2001).
11. Christensen, U. R. Zonal flow driven by strongly supercritical convection in rotating spherical shells. *J. Fluid Mech.* **470**, 115–133 (2002).
12. Yano, J. I., Talagrand, O. & Drossart, P. Origins of atmospheric zonal winds. *Nature* **421**, 36 (2003).
13. Yano, J. I., Talagrand, O. & Drossart, P. Deep two-dimensional turbulence: An idealized model for atmospheric jets of the giant outer planets. *Geophys. Astrophys. Fluid Dyn.* **99**, 137–150 (2005).
14. Aurnou, J. M. & Heimpel, M. H. Zonal jets in rotating convection with mixed mechanical boundary conditions. *Icarus* **169**, 492–498 (2004).
15. Ingersoll, A. P. Atmospheric dynamics of the outer planets. *Science* **248**, 308–315 (1990).
16. Rhines, P. B. Waves and turbulence on a beta-plane. *J. Fluid Mech.* **69**, 417–443 (1975).
17. Schoff, R. & Colin de Verdiere, A. Taylor columns between concentric spheres. *Geophys. Astrophys. Fluid Dyn.* **86**, 43–73 (1997).
18. Busse, F. H. A simple model of convection in the Jovian atmosphere. *Icarus* **20**, 255–260 (1976).
19. Manneville, J. B. & Olson, P. Banded convection in rotating fluid spheres and the circulation of the Jovian atmosphere. *Icarus* **122**, 242–250 (1996).
20. Jones, C. A., Rotvig, J. & Abdulrahman, A. Multiple jets and zonal flow on Jupiter. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL016980 (2003).
21. Read, P. L. *et al.* Jupiter's and Saturn's convectively driven banded jets in the laboratory. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020106 (2004).
22. Atkinson, D. H., Pollack, J. B. & Seiff, A. The Galileo Probe Doppler Wind Experiment: Measurement of the deep zonal winds on Jupiter. *J. Geophys. Res.* **103**, 22911–22928 (1998).
23. Kirk, R. L. & Stevenson, D. J. Hydromagnetic constraints on deep zonal flow in the giant planets. *Astrophys. J.* **316**, 836–846 (1987).
24. Aubert, J., Brito, D., Nataf, H.-C., Cardin, P. & Masson, J. P. A systematic experimental study of spherical shell convection in water and liquid gallium. *Phys. Earth. Planet. Inter.* **128**, 51–74 (2001).
25. Evonuk, M. & Glatzmaier, G. A. 2D studies of various approximations used for modeling convection in the giant planets. *Geophys. Astrophys. Fluid Dyn.* **98**, 241–255 (2004).
26. Hubbard, W. B. Gravitational signature of Jupiter's deep zonal flows. *Icarus* **137**, 357–359 (1999).
27. Martinez, C. NASA selects new frontier mission concept study. (<http://www.jpl.nasa.gov/news/news.cfm?release=2005-090>) (2005).
28. Wicht, J. Inner-core conductivity in numerical dynamo simulations. *Phys. Earth Planet. Inter.* **132**, 281–302 (2002).
29. Al-Shamali, F. M., Heimpel, M. H. & Aurnou, J. M. Varying the spherical shell geometry in rotating thermal convection. *Geophys. Astrophys. Fluid Dyn.* **98**, 153–169 (2004).
30. Kuang, W.-J. & Bloxham, J. Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: Weak and strong field dynamo action. *J. Comp. Phys.* **153**, 51–81 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding was provided by NSERC Canada, UCLA, and the DFG Germany priority programme 'Geomagnetic variations'. Computational resources were provided by the Western Canada Research Grid (West Grid).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (mheimpel@phys.ualberta.ca).

Two-dimensional gas of massless Dirac fermions in graphene

K. S. Novoselov¹, A. K. Geim¹, S. V. Morozov², D. Jiang¹, M. I. Katsnelson³, I. V. Grigorieva¹, S. V. Dubonos² & A. A. Firsov²

Quantum electrodynamics (resulting from the merger of quantum mechanics and relativity theory) has provided a clear understanding of phenomena ranging from particle physics to cosmology and from astrophysics to quantum chemistry^{1–3}. The ideas underlying quantum electrodynamics also influence the theory of condensed matter^{4,5}, but quantum relativistic effects are usually minute in the known experimental systems that can be described accurately by the non-relativistic Schrödinger equation. Here we report an experimental study of a condensed-matter system (graphene, a single atomic layer of carbon^{6,7}) in which electron transport is essentially governed by Dirac's (relativistic) equation. The charge carriers in graphene mimic relativistic particles with zero rest mass and have an effective 'speed of light' $c_* \approx 10^6 \text{ m s}^{-1}$. Our study reveals a variety of unusual phenomena that are characteristic of two-dimensional Dirac fermions. In particular we have observed the following: first, graphene's conductivity never falls below a minimum value corresponding to the quantum unit of conductance, even when concentrations of charge carriers tend to zero; second, the integer quantum Hall effect in graphene is anomalous in that it occurs at half-integer filling factors; and third, the cyclotron mass m_c of massless carriers in graphene is described by $E = m_c c_*^2$. This two-dimensional system is not only interesting in itself but also allows access to the subtle and rich physics of quantum electrodynamics in a bench-top experiment.

Graphene is a monolayer of carbon atoms packed into a dense honeycomb crystal structure that can be viewed as an individual atomic plane extracted from graphite, as unrolled single-wall carbon nanotubes or as a giant flat fullerene molecule. This material has not been studied experimentally before and, until recently^{6,7}, was presumed not to exist in the free state. To obtain graphene samples we used the original procedures described in ref. 6, which involve the micromechanical cleavage of graphite followed by the identification and selection of monolayers by using a combination of optical microscopy, scanning electron microscopy and atomic-force microscopy. The selected graphene films were further processed into multi-terminal devices such as that shown in Fig. 1, by following standard microfabrication procedures⁷. Despite being only one atom thick and unprotected from the environment, our graphene devices remain stable under ambient conditions and exhibit high mobility of charge carriers. Below we focus on the physics of 'ideal' (single-layer) graphene, which has a different electronic structure and exhibits properties qualitatively different from those characteristic of either ultrathin graphite films (which are semimetals whose material properties were studied recently^{7–10}) or even of other devices consisting of just two layers of graphene (see below).

Figure 1 shows the electric field effect^{7–9} in graphene. Its conductivity σ increases linearly with increasing gate voltage V_g for both polarities, and the Hall effect changes its sign at $V_g \approx 0$. This

behaviour shows that substantial concentrations of electrons (holes) are induced by positive (negative) gate voltages. Away from the transition region $V_g \approx 0$, Hall coefficient $R_H = 1/ne$ varies as $1/V_g$, where n is the concentration of electrons or holes and e is the electron charge. The linear dependence $1/R_H \propto V_g$ yields $n = \alpha V_g$ with $\alpha \approx 7.3 \times 10^{10} \text{ cm}^{-2} \text{ V}^{-1}$, in agreement with the theoretical estimate $n/V_g \approx 7.2 \times 10^{10} \text{ cm}^{-2} \text{ V}^{-1}$ for the surface charge density induced by the field effect (see the caption to Fig. 1). The agreement indicates that all the induced carriers are mobile and that there are no trapped charges in graphene. From the linear dependence $\sigma(V_g)$ we found carrier mobilities $\mu = \sigma/ne$, which reached $15,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for both electrons and holes, were independent of temperature T between 10 and 100 K and were probably still limited by defects in parent graphite.

To characterize graphene further, we studied Shubnikov-de Haas oscillations (SdHOs). Figure 2 shows examples of these oscillations for different magnetic fields B , gate voltages and temperatures. Unlike ultrathin graphite⁷, graphene exhibits only one set of SdHO for both electrons and holes. By using standard fan diagrams^{7,8} we have determined the fundamental SdHO frequency B_F for various V_g . The resulting dependence of B_F on n is plotted in Fig. 3a. Both carriers exhibit the same linear dependence $B_F = \beta n$, with $\beta \approx 1.04 \times 10^{-15} \text{ T m}^2$ ($\pm 2\%$). Theoretically, for any two-dimensional (2D) system β is defined only by its degeneracy f so that $B_F = \phi_0 n/f$, where $\phi_0 = 4.14 \times 10^{-15} \text{ T m}^2$ is the flux quantum. Comparison with the experiment yields $f = 4$, in agreement with the double-spin and double-valley degeneracy expected for graphene^{11,12} (see caption to Fig. 2). Note, however, an anomalous feature of SdHO in graphene, which is their phase. In contrast to conventional metals, graphene's longitudinal resistance $\rho_{xx}(B)$ exhibits maxima rather than minima at integer values of the Landau filling factor ν (Fig. 2a). Figure 3b emphasizes this fact by comparing the phase of SdHO in graphene with that in a thin graphite film⁷. The origin of the 'odd' phase is explained below.

Another unusual feature of 2D transport in graphene clearly reveals itself in the dependence of SdHO on T (Fig. 2b). Indeed, with increasing T the oscillations at high V_g (high n) decay more rapidly. One can see that the last oscillation ($V_g \approx 100 \text{ V}$) becomes practically invisible at 80 K, whereas the first one ($V_g < 10 \text{ V}$) clearly survives at 140 K and remains notable even at room temperature. To quantify this behaviour we measured the T -dependence of SdHO's amplitude at various gate voltages and magnetic fields. The results could be fitted accurately (Fig. 3c) by the standard expression $T/\sinh(2\pi^2 k_B T m_c / \hbar e B)$, which yielded m_c varying between ~ 0.02 and $0.07 m_0$ (m_0 is the free electron mass). Changes in m_c are well described by a square-root dependence $m_c \propto n^{1/2}$ (Fig. 3d).

To explain the observed behaviour of m_c , we refer to the semiclassical expressions $B_F = (\hbar/2\pi e)S(E)$ and $m_c = (\hbar^2/2\pi)\partial S(E)/\partial E$,

¹Manchester Centre for Mesoscience and Nanotechnology, University of Manchester, Manchester M13 9PL, UK. ²Institute for Microelectronics Technology, 142432 Chernogolovka, Russia. ³Institute for Molecules and Materials, Radboud University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

where $S(E) = \pi k^2$ is the area in k -space of the orbits at the Fermi energy $E(k)$ (ref. 13). If these expressions are combined with the experimentally found dependences $m_c \propto n^{1/2}$ and $B_F = (h/4e)n$ it is straightforward to show that S must be proportional to E^2 , which yields $E \propto k$. The data in Fig. 3 therefore unambiguously prove the linear dispersion $E = \hbar k c^*$ for both electrons and holes with a common origin at $E = 0$ (refs 11, 12). Furthermore, the above equations also imply $m_c = E/c^* = (\hbar^2 n / 4\pi c^*)^{1/2}$ and the best fit to our data yields $c^* \approx 10^6 \text{ m s}^{-1}$, in agreement with band structure calculations^{11,12}. The semiclassical model employed is fully justified by a recent theory for graphene¹⁴, which shows that SdHO's amplitude can indeed be described by the above expression $T/\sinh(2\pi^2 k_B T m_c / \hbar e B)$ with $m_c = E/c^*$. Therefore, even though the linear spectrum of fermions in graphene (Fig. 3e) implies zero rest mass, their cyclotron mass is not zero.

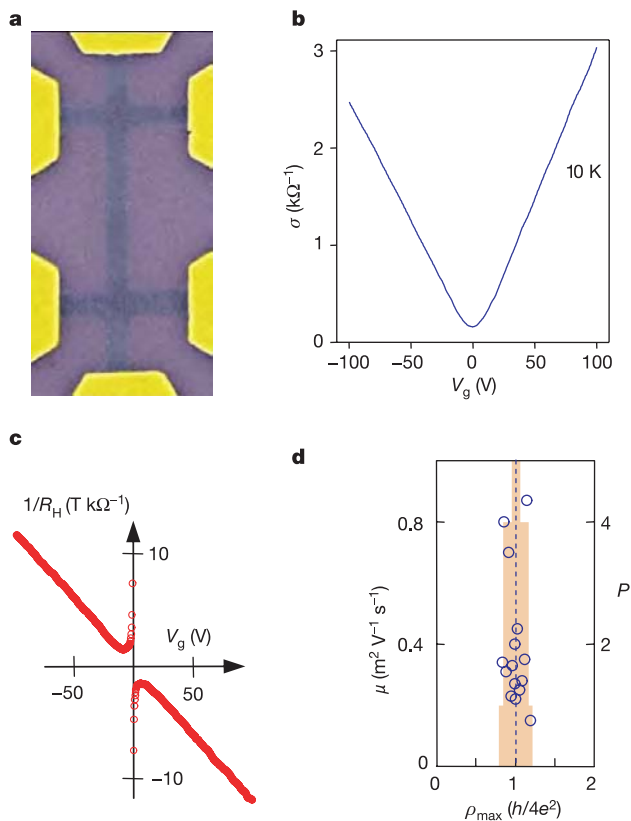


Figure 1 | Electric field effect in graphene. **a**, Scanning electron microscope image of one of our experimental devices (the width of the central wire is $0.2 \mu\text{m}$). False colours are chosen to match real colours as seen in an optical microscope for large areas of the same material. **b**, **c**, Changes in graphene's conductivity σ (**b**) and Hall coefficient R_H (**c**) as a function of gate voltage V_g . σ and R_H were measured in magnetic fields B of 0 and 2 T, respectively. The induced carrier concentrations n are described in ref. 7; $n/V_g = \epsilon_0 \epsilon / te$, where ϵ_0 and ϵ are the permittivities of free space and SiO_2 , respectively, and $t \approx 300 \text{ nm}$ is the thickness of SiO_2 on top of the Si wafer used as a substrate. $R_H = 1/ne$ is inverted to emphasize the linear dependence $n \propto V_g$. $1/R_H$ diverges at small n because the Hall effect changes its sign at about $V_g = 0$, indicating a transition between electrons and holes. Note that the transition region ($R_H \approx 0$) was often shifted from zero V_g as a result of chemical doping⁷, but annealing of our devices in vacuum normally allowed us to eliminate the shift. The extrapolation of the linear slopes $\sigma(V_g)$ for electrons and holes results in their intersection at a value of σ indistinguishable from zero. **d**, Maximum values of resistivity $\rho = 1/\sigma$ (circles) exhibited by devices with different mobilities μ (left y axis). The histogram (orange background) shows the number P of devices exhibiting ρ_{max} within 10% intervals around the average value of $\sim h/4e^2$. Several of the devices shown were made from two or three layers of graphene, indicating that the quantized minimum conductivity is a robust effect and does not require 'ideal' graphene.

The unusual response of massless fermions to a magnetic field is highlighted further by their behaviour in the high-field limit, at which SdHOs evolve into the quantum Hall effect (QHE). Figure 4 shows the Hall conductivity σ_{xy} of graphene plotted as a function of electron and hole concentrations in a constant B . Pronounced QHE plateaux are visible, but they do not occur in the expected sequence $\sigma_{xy} = (4e^2/h)N$, where N is integer. On the contrary, the plateaux correspond to half-integer ν so that the first plateau occurs at $2e^2/h$ and the sequence is $(4e^2/h)(N + 1/2)$. The transition from the lowest hole ($\nu = -1/2$) to the lowest electron ($\nu = +1/2$) Landau level (LL) in graphene requires the same number of carriers ($\Delta n = 4B/\phi_0 \approx 1.2 \times 10^{12} \text{ cm}^{-2}$) as the transition between other nearest levels (compare the distances between minima in ρ_{xx}). This results in a ladder of equidistant steps in σ_{xy} that are not interrupted when passing through zero. To emphasize this highly unusual behaviour, Fig. 4 also shows σ_{xy} for a graphite film consisting of only two graphene layers, in which the sequence of plateaux returns to normal and the first plateau is at $4e^2/h$, as in the conventional QHE. We attribute this qualitative transition between graphene and its two-layer counterpart to the fact that fermions in the latter exhibit a finite mass near $n \approx 0$ and can no longer be described as massless Dirac particles.

The half-integer QHE in graphene has recently been suggested by two theory groups^{15,16}, stimulated by our work on thin graphite films⁷ but unaware of the present experiment. The effect is single-particle and is intimately related to subtle properties of massless Dirac fermions, in particular to the existence of both electron-like and hole-like Landau states at exactly zero energy^{14–17}. The latter can be viewed as a direct consequence of the Atiyah–Singer index theorem that is important in quantum field theory and the theory of superstrings^{18,19}. For 2D massless Dirac fermions, the theorem guarantees the existence of Landau states at $E = 0$ by relating the difference in the number of such states with opposite chiralities to the total flux through the system (magnetic field can be inhomogeneous).

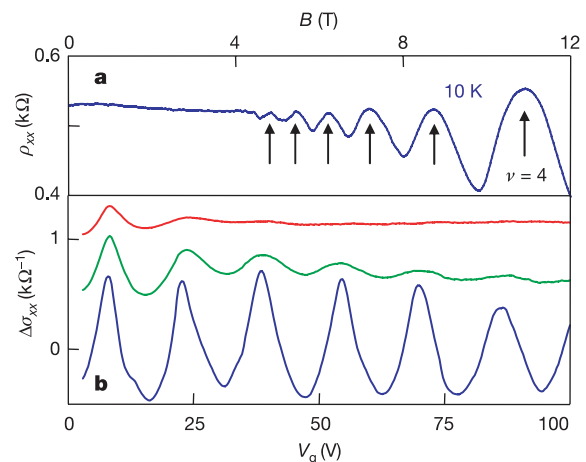


Figure 2 | Quantum oscillations in graphene. SdHO at constant gate voltage $V_g = -60 \text{ V}$ as a function of magnetic field B (**a**) and at constant $B = 12 \text{ T}$ as a function of V_g (**b**). Because μ does not change greatly with V_g , the measurements at constant B (at a constant $\omega_c \tau = \mu B$) were found more informative. In **b**, SdHOs in graphene are more sensitive to T at high carrier concentrations: blue, $T = 20 \text{ K}$; green, $T = 80 \text{ K}$; red, $T = 140 \text{ K}$. The $\Delta\sigma_{xx}$ curves were obtained by subtracting a smooth (nearly linear) increase in σ with increasing V_g and are shifted for clarity. SdHO periodicity ΔV_g at constant B is determined by the density of states at each Landau level ($\alpha \Delta V_g = fB/\phi_0$), which for the observed periodicity of $\sim 15.8 \text{ V}$ at $B = 12 \text{ T}$ yields a quadruple degeneracy. Arrows in **a** indicate integer ν (for example, $\nu = 4$ corresponds to 10.9 T) as found from SdHO frequency $B_F \approx 43.5 \text{ T}$. Note the absence of any significant contribution of universal conductance fluctuations (see also Fig. 1) and weak localization magnetoresistance, which are normally intrinsic for 2D materials with so high resistivity.

To explain the half-integer QHE qualitatively, we invoke the formal expression^{2,14–17} for the energy of massless relativistic fermions in quantized fields, $E_N = [2\hbar c^2 B(N + 1/2 \pm 1/2)]^{1/2}$. In quantum electrodynamics, the sign $\pm$ describes two spins, whereas in graphene it refers to ‘pseudospins’. The latter have nothing to do with the real spin but are ‘built in’ to the Dirac-like spectrum of graphene; their origin can be traced to the presence of two carbon sublattices. The above formula shows that the lowest LL ($N = 0$) appears at $E = 0$ (in agreement with the index theorem) and accommodates fermions with only one (minus) projection of the pseudospin. All other levels $N \geq 1$ are occupied by fermions with both ($\pm$) pseudospins. This implies that for $N = 0$ the degeneracy is half of that for any other N . Alternatively, one can say that all LLs have the same ‘compound’ degeneracy but the zero-energy LL is shared equally by electrons and holes. As a result the first Hall plateau occurs at half the normal filling and, oddly, both $\nu = -1/2$ and $+1/2$ correspond to the same LL ($N = 0$). All other levels have normal degeneracy $4B/\phi_0$ and therefore remain shifted by the same $1/2$ from the standard sequence. This explains the QHE at $\nu = N + 1/2$ and, at the same time, the ‘odd’ phase of SdHO (minima in ρ_{xx} correspond to plateaux in ρ_{xy} and therefore occur at half-integer ν ; see Figs 2 and 4), in agreement with theory^{14–17}. Note, however, that from another perspective the phase shift can be viewed as the direct manifestation of Berry’s phase acquired by Dirac fermions moving in magnetic field^{20,21}.

Finally, we return to zero-field behaviour and discuss another feature related to graphene’s relativistic-like spectrum. The spectrum implies vanishing concentrations of both carriers near the Dirac point $E = 0$ (Fig. 3e), which suggests that low- T resistivity of the

zero-gap semiconductor should diverge at $V_g \approx 0$. However, neither of our devices showed such behaviour. On the contrary, in the transition region between holes and electrons graphene’s conductivity never falls below a well-defined value, practically independent of T between 4 K and 100 K. Figure 1c plots values of the maximum resistivity $\rho_{\max}$ found in 15 different devices at zero B , which within an experimental error of $\sim 15\%$ all exhibit $\rho_{\max} \approx 6.5 \text{ k}\Omega$ independently of their mobility, which varies by a factor of 10. Given the quadruple degeneracy f , it is obvious to associate $\rho_{\max}$ with $h/fe^2 = 6.45 \text{ k}\Omega$, where h/e^2 is the resistance quantum. We emphasize that it is the resistivity (or conductivity) rather than the resistance (or conductance) that is quantized in graphene (that is, resistance R measured experimentally scaled in the usual manner as $R = \rho L/w$ with changing length L and width w of our devices). Thus, the effect is completely different from the conductance quantization observed previously in quantum transport experiments.

However surprising it may be, the minimum conductivity is an intrinsic property of electronic systems described by the Dirac equation^{22–25}. It is due to the fact that, in the presence of disorder, localization effects in such systems are strongly suppressed and emerge only at exponentially large length scales. Assuming the absence of localization, the observed minimum conductivity can be explained qualitatively by invoking Mott’s argument²⁶ that the mean free path l of charge carriers in a metal can never be shorter than their wavelength λ_F . Then, $\sigma = ne\mu$ can be rewritten as $\sigma = (e^2/h)k_F l$, so σ cannot be smaller than $\sim e^2/h$ for each type of carrier. This argument is known to have failed for 2D systems with a parabolic spectrum in which disorder leads to localization and eventually to insulating behaviour^{22,23}. For 2D Dirac fermions, no localization is expected^{22–25} and, accordingly, Mott’s argument can be used. Although there is a broad theoretical consensus^{15,16,23–28} that a 2D gas of Dirac fermions should exhibit a minimum conductivity of about e^2/h , this quantization was not expected to be accurate and most theories suggest a value of $\sim e^2/\pi h$, in disagreement with the experiment.

Thus, graphene exhibits electronic properties that are distinctive for a 2D gas of particles described by the Dirac equation rather than the Schrödinger equation. The work shows a possibility of studying

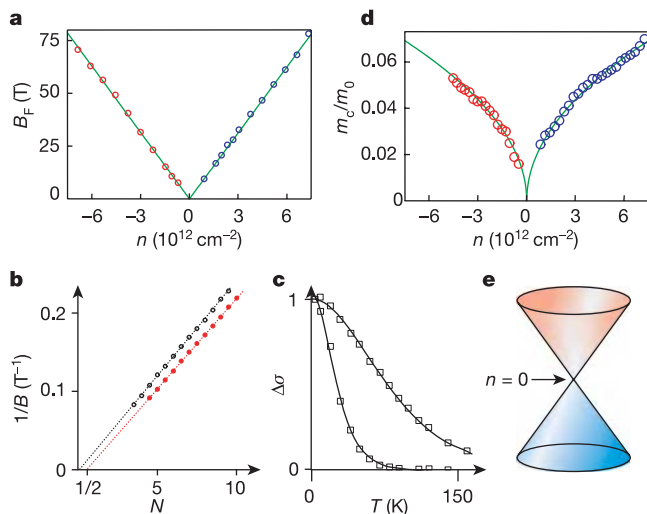


Figure 3 | Dirac fermions of graphene. **a**, Dependence of B_F on carrier concentration n (positive n corresponds to electrons; negative to holes). **b**, Examples of fan diagrams used in our analysis⁷ to find B_F . N is the number associated with different minima of oscillations. The lower and upper curves are for graphene (sample of Fig. 2a) and a 5-nm-thick film of graphite with a similar value of B_F , respectively. Note that the curves extrapolate to different origins, namely to $N = 1/2$ and $N = 0$. In graphene, curves for all n extrapolate to $N = 1/2$ (compare ref. 7). This indicates a phase shift of π with respect to the conventional Landau quantization in metals. The shift is due to Berry’s phase^{14,20}. **c**, Examples of the behaviour of SdHO amplitude $\Delta\sigma$ (symbols) as a function of T for $m_c \approx 0.069$ and $0.023m_0$ (see the dependences showing the rapid and slower decay with increasing T , respectively); solid curves are best fits. **d**, Cyclotron mass m_c of electrons and holes as a function of their concentration. Symbols are experimental data, solid curves the best fit to theory. **e**, Electronic spectrum of graphene, as inferred experimentally and in agreement with theory. This is the spectrum of a zero-gap 2D semiconductor that describes massless Dirac fermions with $c = 1/300$ the speed of light.

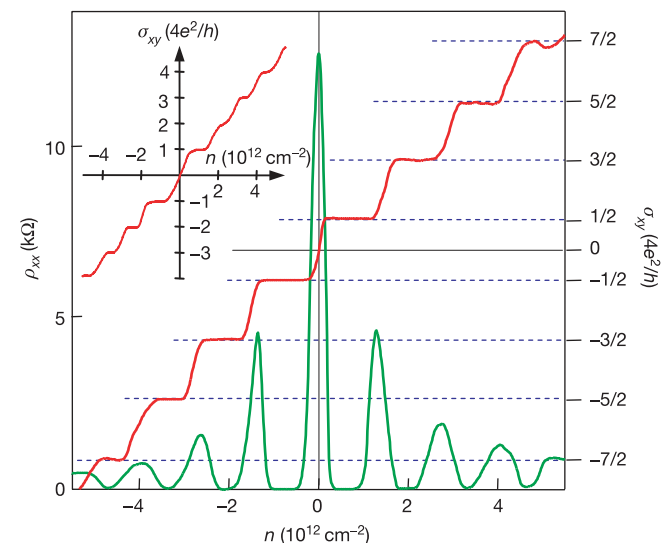


Figure 4 | QHE for massless Dirac fermions. Hall conductivity σ_{xy} and longitudinal resistivity ρ_{xx} of graphene as a function of their concentration at $B = 14 \text{ T}$ and $T = 4 \text{ K}$. $\sigma_{xy} \equiv (4e^2/h)\nu$ is calculated from the measured dependences of $\rho_{xy}(V_g)$ and $\rho_{xx}(V_g)$ as $\sigma_{xy} = \rho_{xy}/(\rho_{xy}^2 + \rho_{xx}^2)$. The behaviour of $1/\rho_{xy}$ is similar but exhibits a discontinuity at $V_g \approx 0$, which is avoided by plotting σ_{xy} . Inset: σ_{xy} in ‘two-layer graphene’ where the quantization sequence is normal and occurs at integer ν . The latter shows that the half-integer QHE is exclusive to ‘ideal’ graphene.

phenomena of the quantum field theory in a condensed-matter experiment.

Received 14 July; accepted 12 September 2005.

1. Rose, M. E. *Relativistic Electron Theory* (Wiley, New York, 1961).
2. Berestetskii, V. B., Lifshitz, E. M. & Pitaevskii, L. P. *Relativistic Quantum Theory* (Pergamon, Oxford, 1971).
3. Lai, D. Matter in strong magnetic fields. *Rev. Mod. Phys.* **73**, 629–662 (2001).
4. Fradkin, E. *Field Theories of Condensed Matter Systems* (Westview, Oxford, 1997).
5. Volovik, G. E. *The Universe in a Helium Droplet* (Clarendon, Oxford, 2003).
6. Novoselov, K. S. *et al.* Two dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).
7. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
8. Zhang, Y., Small, J. P., Amori, M. E. S. & Kim, P. Electric field modulation of galvanomagnetic properties of mesoscopic graphite. *Phys. Rev. Lett.* **94**, 176803 (2005).
9. Berger, C. *et al.* Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics. *J. Phys. Chem. B* **108**, 19912–19916 (2004).
10. Bunch, J. S., Yaish, Y., Brink, M., Bolotin, K. & McEuen, P. L. Coulomb oscillations and Hall effect in quasi-2D graphite quantum dots. *Nano Lett.* **5**, 287–290 (2005).
11. Dresselhaus, M. S. & Dresselhaus, G. Intercalation compounds of graphite. *Adv. In Phys.* **51**, 1–186 (2002).
12. Brandt, N. B., Chudinov, S. M. & Ponomarev, Y. G. *Semimetals 1: Graphite and Its Compounds* (North-Holland, Amsterdam, 1988).
13. Vonsovsky, S. V. & Katsnelson, M. I. *Quantum Solid State Physics* (Springer, New York, 1989).
14. Gusynin, V. P. & Sharapov, S. G. Magnetic oscillations in planar systems with the Dirac-like spectrum of quasiparticle excitations. *Phys. Rev. B* **71**, 125124 (2005).
15. Gusynin, V. P. & Sharapov, S. G. Unconventional integer quantum Hall effect in grapheme. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0506575>) (2005).
16. Peres, N. M. R., Guinea, F. & Castro Neto, A. H. Electronic properties of two-dimensional carbon. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0506709>) (2005).
17. Zheng, Y. & Ando, T. Hall conductivity of a two-dimensional graphite system. *Phys. Rev. B* **65**, 245420 (2002).
18. Kaku, M. *Introduction to Superstrings* (Springer, New York, 1988).
19. Nakahara, M. *Geometry, Topology and Physics* (IOP, Bristol, 1990).
20. Mikitik, G. P. & Sharlai, Yu. V. Manifestation of Berry's phase in metal physics. *Phys. Rev. Lett.* **82**, 2147–2150 (1999).
21. Luk'yanchuk, I. A. & Kopelevich, Y. Phase analysis of quantum oscillations in graphite. *Phys. Rev. Lett.* **93**, 166402 (2004).
22. Abrahams, E., Anderson, P. W., Licciardello, D. C. & Ramakrishnan, T. V. Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.* **42**, 673–676 (1979).
23. Fradkin, E. Critical behaviour of disordered degenerate semiconductors. *Phys. Rev. B* **33**, 3263–3268 (1986).
24. Lee, P. A. Localized states in a d-wave superconductor. *Phys. Rev. Lett.* **71**, 1887–1890 (1993).
25. Ziegler, K. Delocalization of 2D Dirac fermions: The role of a broken symmetry. *Phys. Rev. Lett.* **80**, 3113–3116 (1998).
26. Mott, N. F. & Davis, E. A. *Electron Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1979).
27. Morita, Y. & Hatsugai, Y. Near critical states of random Dirac fermions. *Phys. Rev. Lett.* **79**, 3728–3731 (1997).
28. Nersisyan, A. A., Tsvetlik, A. M. & Wenger, F. Disorder effects in two-dimensional d-wave superconductors. *Phys. Rev. Lett.* **72**, 2628–2631 (1997).

Acknowledgements We thank L. Glazman, V. Falko, S. Sharapov and A. Castro Neto for discussions. K.S.N. was supported by Leverhulme Trust. S.V.M., S.V.D. and A.A.F. acknowledge support from the Russian Academy of Science and INTAS. This research was funded by the EPSRC (UK).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K.G. (geim@man.ac.uk) or K.S.N. (kostya@man.ac.uk).

Experimental observation of the quantum Hall effect and Berry's phase in graphene

Yuanbo Zhang¹, Yan-Wen Tan¹, Horst L. Stormer^{1,2} & Philip Kim¹

When electrons are confined in two-dimensional materials, quantum-mechanically enhanced transport phenomena such as the quantum Hall effect can be observed. Graphene, consisting of an isolated single atomic layer of graphite, is an ideal realization of such a two-dimensional system. However, its behaviour is expected to differ markedly from the well-studied case of quantum wells in conventional semiconductor interfaces. This difference arises from the unique electronic properties of graphene, which exhibits electron–hole degeneracy and vanishing carrier mass near the point of charge neutrality^{1,2}. Indeed, a distinctive half-integer quantum Hall effect has been predicted^{3–5} theoretically, as has the existence of a non-zero Berry's phase (a geometric quantum phase) of the electron wavefunction—a consequence of the exceptional topology of the graphene band structure^{6,7}. Recent advances in micromechanical extraction and fabrication techniques for graphite structures^{8–12} now permit such exotic two-dimensional electron systems to be probed experimentally. Here we report an experimental investigation of magneto-transport in a high-mobility single layer of graphene. Adjusting the chemical potential with the use of the electric field effect, we observe an unusual half-integer quantum Hall effect for both electron and hole carriers in graphene. The relevance of Berry's phase to these experiments is confirmed by magneto-oscillations. In addition to their purely scientific interest, these unusual quantum transport phenomena may lead to new applications in carbon-based electronic and magneto-electronic devices.

The low-energy band structure of graphene can be approximated as cones located at two inequivalent Brillouin zone corners (Fig. 1a, left inset). In these cones, the two-dimensional (2D) energy dispersion relation is linear and the electron dynamics can be treated as 'relativistic', in which the Fermi velocity v_F of the graphene substitutes for the speed of light. In particular, at the apex of the cones (termed the Dirac point), electrons and holes (particles and antiparticles) are degenerate. Landau-level (LL) formation for electrons in this system under a perpendicular magnetic field, B , has been studied theoretically using an analogy to 2 + 1-dimensional quantum electrodynamics^{2,3}, in which the Landau level energy is given by

$$E_n = \text{sgn}(n) \sqrt{2e\hbar v_F^2 |n| B} \quad (1)$$

Here e and $\hbar = h/2\pi$ are electron charge and Planck's constant divided by 2π , and the integer n represents an electron-like ($n > 0$) or a hole-like ($n < 0$) LL index. Crucially, a single LL with $n = 0$ and $E_0 = 0$ also occurs. When only low-lying LLs ($|n| < 10^4$ for $B = 10$ T) are occupied, the separation of E_n is much larger than the Zeeman spin splitting, so each LL has a degeneracy $g_s = 4$, accounting for spin degeneracy and sublattice degeneracy. Previous studies of mesoscopic graphite samples consisting of a few layers of graphene exhibited magneto-oscillations associated with the LL formation by

electron-like and hole-like carriers tuned by the electric field effect^{8,9,11}. However, the quantum Hall effect (QHE) was not observed in these samples, possibly as a result of their low mobility and/or the residual three-dimensional nature of the specimens.

The high-mobility graphene samples used in our experiments were extracted from Kish graphite (Toshiba Ceramics) on degenerately doped Si wafers with a 300-nm SiO₂ coating layer, by using micromechanical manipulation similar to that described in ref. 8.

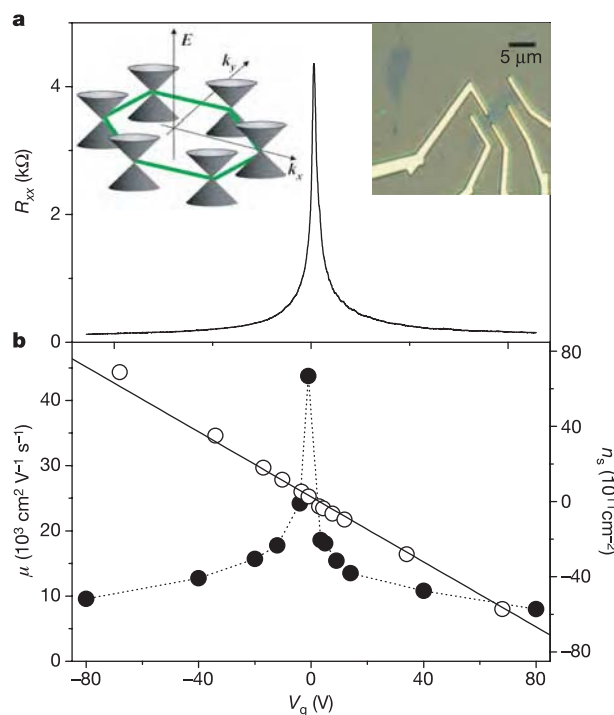


Figure 1 | Resistance, carrier density, and mobility of graphene measured at 1.7 K at different gate voltages. **a**, Changes in resistance as a function of gate voltage in a graphene device shown in the optical microscope image in the right inset. The position of the resistance peaks varies from device to device, but the peak values are always of the order of 4 kΩ, suggesting a potential quantum-mechanical origin. The left inset shows a schematic diagram of the low-energy dispersion relation near the Dirac points in the graphene Brillouin zone. Only two Dirac cones are nonequivalent to each other, producing a twofold valley degeneracy in the band structure. **b**, Charge carrier density (open circles) and mobility (filled circles) of graphene as a function of gate voltage. The solid line corresponds to the estimated charge induced by the gate voltage, $n_s = C_g V_g / e$, assuming a gate capacitance C_g of $115 \text{ aF } \mu\text{m}^{-2}$ obtained from geometrical considerations.

¹Department of Physics, ²Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA.

Interference-induced colour shifts, cross-correlated with an atomic force microscopy profile, allow us to identify the number of deposited graphene layers from optical images of the samples (Supplementary Information). After a suitable graphene sample has been selected, electron beam lithography followed by thermally evaporated Au/Cr (30 nm and 5 nm, respectively) defines multiple electrodes for transport measurement (Fig. 1a, right inset). With the use of a Hall-bar-type electrode configuration, the magnetoresistance R_{xx} and Hall resistance R_{xy} are measured. Applying a gate voltage, V_g , to the Si substrate controls the charge density in the graphene samples.

Figure 1a shows the gate modulation of R_{xx} at zero magnetic field in a typical graphene device whose lateral size is $\sim 3 \mu\text{m}$. Whereas R_{xx} remains in the $\sim 100\text{-}\Omega$ range at high carrier density, a sharp peak at $\sim 4\text{ k}\Omega$ is observed at $V_g \approx 0$. Although different samples show slightly different peak values and peak positions, similar behaviours were observed in three other graphene samples that we measured. The existence of this sharp peak is consistent with the reduced carrier density as E_F approaches the Dirac point of graphene, at which the density of states vanishes. Thus, the gate voltage corresponding to the charge-neutral Dirac point, V_{Dirac} , can be determined from this peak position. A separate Hall measurement provides a measure for the sheet carrier density, n_s , and for the mobility, μ , of the sample, as shown in Fig. 1b, assuming a simple Drude model. The sign of n_s changes at $V_g = V_{\text{Dirac}}$, indicating that E_F does indeed cross the charge-neutral point. Mobilities are higher than $10^4\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ for the entire gate voltage range, considerably exceeding the quality of graphene samples studied previously^{8,9}.

The exceptionally high-mobility graphene samples allow us to

investigate transport phenomena in the extreme magnetic quantum limit, such as the QHE. Figure 2a shows R_{xy} and R_{xx} for the sample of Fig. 1 as a function of magnetic field B at a fixed gate voltage $V_g > V_{\text{Dirac}}$. The overall positive R_{xy} indicates that the contribution is mainly from electrons. At high magnetic field, $R_{xy}(B)$ exhibits plateaux and R_{xx} is vanishing, which are the hallmark of the QHE. At least two well-defined plateaux with values $(2e^2/h)^{-1}$ and $(6e^2/h)^{-1}$, followed by a developing $(10e^2/h)^{-1}$ plateau, are observed before the QHE features transform into Shubnikov de Haas (SdH) oscillations at lower magnetic field. The quantization of R_{xy} for these first two plateaux is better than 1 part in 10^4 , precise within the instrumental uncertainty. We observed the equivalent QHE features for holes with negative R_{xy} values (Fig. 2a, inset). Alternatively, we can probe the QHE in both electrons and holes by fixing the magnetic field and changing V_g across the Dirac point. In this case, as V_g increases, first holes ($V_g < V_{\text{Dirac}}$) and later electrons ($V_g > V_{\text{Dirac}}$) fill successive Landau levels and exhibit the QHE. This yields an antisymmetric (symmetric) pattern of R_{xy} (R_{xx}) in Fig. 2b, with R_{xy} quantization in accordance with

$$R_{xy}^{-1} = \pm g_s(n + 1/2)e^2/h \quad (2)$$

where n is a non-negative integer and $\pm$ stands for electrons and holes, respectively. This quantization condition can be translated to the quantized filling factor $\nu = \pm g_s(n + 1/2)$ in the usual QHE language. In addition, there is an oscillatory structure developed near the Dirac point. Although this structure is reproducible for any given sample, its shape varies from device to device, suggesting potentially mesoscopic effects depending on the details of the sample geometry¹³. Although the QHE has been observed in many 2D

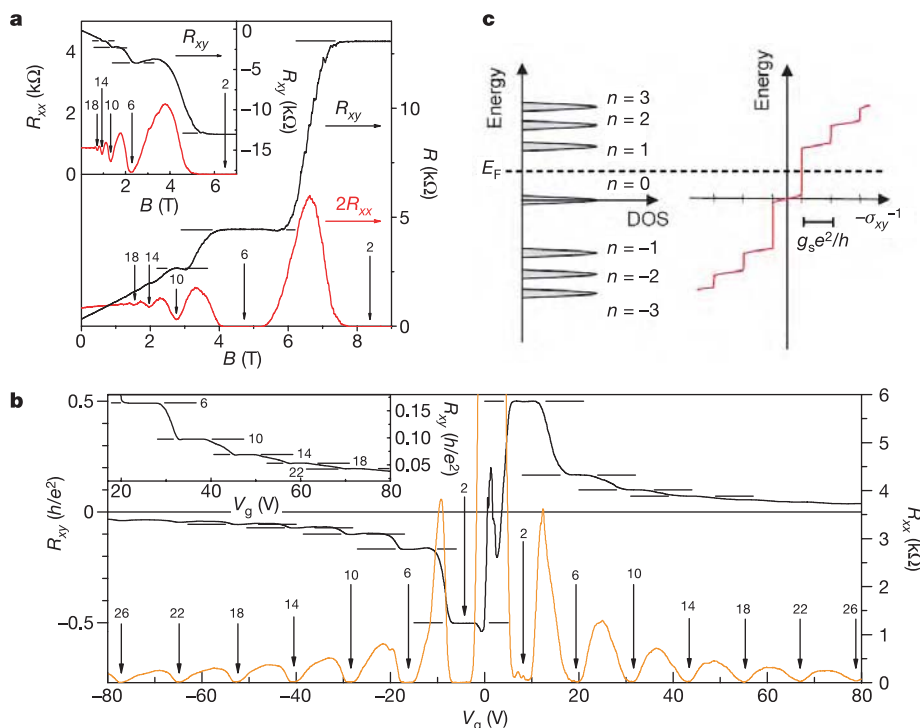


Figure 2 | Quantized magnetoresistance and Hall resistance of a graphene device. **a**, Hall resistance (black) and magnetoresistance (red) measured in the device in Fig. 1 at $T = 30\text{ mK}$ and $V_g = 15\text{ V}$. The vertical arrows and the numbers on them indicate the values of B and the corresponding filling factor ν of the quantum Hall states. The horizontal lines correspond to $h/e^2\nu$ values. The QHE in the electron gas is shown by at least two quantized plateaux in R_{xy} , with vanishing R_{xx} in the corresponding magnetic field regime. The inset shows the QHE for a hole gas at $V_g = -4\text{ V}$, measured at 1.6 K . The quantized plateau for filling factor $\nu = 2$ is well defined, and the second and third plateaux with $\nu = 6$ and $\nu = 10$ are also resolved. **b**, Hall

resistance (black) and magnetoresistance (orange) as a function of gate voltage at fixed magnetic field $B = 9\text{ T}$, measured at 1.6 K . The same convention as in **a** is used here. The upper inset shows a detailed view of high-filling-factor plateaux measured at 30 mK . **c**, A schematic diagram of the Landau level density of states (DOS) and corresponding quantum Hall conductance (σ_{xy}) as a function of energy. Note that, in the quantum Hall states, $\sigma_{xy} = -R_{xy}^{-1}$. The LL index n is shown next to the DOS peak. In our experiment the Fermi energy E_F can be adjusted by the gate voltage, and R_{xy}^{-1} changes by an amount $g_s e^2/h$ as E_F crosses a LL.

systems, the QHE observed in graphene is distinctively different from those 'conventional' QHEs because the quantization condition (equation (2)) is shifted by a half-integer. These unusual quantization conditions are a result of the topologically exceptional electronic structure of graphene, which we discuss below.

The sequence of half-integer multiples of quantum Hall plateaux has been predicted by several theories that combine 'relativistic' Landau levels with the particle-hole symmetry of graphene^{3–5}. This can be easily understood from the calculated LL spectrum (equation (1)) as shown in Fig. 2c. Here we plot the density of states of the g_s -fold degenerate (spin and sublattice) of LLs and the corresponding Hall conductance ($\sigma_{xy} = -R_{xy}^{-1}$, for $R_{xx} \rightarrow 0$) in the quantum Hall regime as a function of energy. σ_{xy} exhibits QHE plateaux when E_F (tuned by V_g) falls between LLs, and jumps by an amount of $g_s e^2/h$ when E_F crosses a LL. Time-reversal invariance guarantees particle-hole symmetry; σ_{xy} is therefore an odd function in energy across the Dirac point². However, in graphene, the $n = 0$ LL is robust—that is, $E_0 = 0$ regardless of the magnetic field—provided that the sublattice symmetry is preserved². Thus, the first plateau of R_{xy}^{-1} for electron and hole is situated exactly at $\pm g_s e^2/2h$. As E_F crosses the next electron (hole) LL, R_{xy}^{-1} increases (decreases) by an amount $g_s e^2/h$, which yields the quantization condition in equation (2).

As noted by several workers, a consequence of the combination of time-reversal symmetry with the novel Dirac point structure can be viewed in terms of Berry's phase arising from the band degeneracy point^{7,14}. A direct implication of Berry's phase in graphene is discussed in the context of the quantum phase of a spin-1/2 pseudo-spinor that describes the sublattice symmetry^{6,15}. This phase is already implicit in the half-integer-shifted quantization rules of the QHE. It can further be probed in the magnetic field regime, in which a semi-classical magneto-oscillation description holds^{16,17}:

$$\Delta R_{xx} = R(B, T) \cos[2\pi(B_F/B + 1/2 + \beta)] \quad (3)$$

Here $R(B, T)$ is the SdH oscillation amplitude, B_F is the frequency of the SdH oscillation in $1/B$, and β is the associated Berry's phase, in the range $0 < \beta < 1$. Berry's phase $\beta = 0$ (or, equivalently, $\beta = 1$) corresponds to the trivial case. A deviation from this value is indicative of new physics with $\beta = 1/2$, implying the existence of Dirac particles⁷. Experimentally, this phase shift in the semi-classical regime can be obtained from an analysis of the SdH fan diagram, in

which the sequence of values of $1/B_n$ of the n th minimum in R_{xx} are plotted against their index n (Fig. 3b). The intercept of linear fit to the data with the n -index axis yields Berry's phase, modulo an integer. The resulting β is very close to 0.5 (Fig. 3b, upper inset), providing further manifestation of the existence of a non-zero Berry's phase in graphene and the presence of Dirac particles. Such a non-zero Berry's phase was not observed in the previous few-layer graphite specimens^{8,11,18}, although there have been claims of hints of a phase shift in earlier measurements on bulk graphite¹⁷. Our data for graphene provide indisputable evidence for such an effect in a solid-state system.

The non-zero Berry's phase observed in the SdH fan diagram is related to the vanishing mass at the Dirac point. We can extract this effective carrier mass m_c from the temperature dependence of the well-developed SdH oscillations at low B field (Fig. 3a, left inset) by using the standard SdH formalism¹⁹. Indeed, the analysis at different gate voltages yields a strong suppression of m_c near the Dirac point. Whereas the high-density ($n_s \sim 5 \times 10^{12} \text{ cm}^{-2}$) carrier gas shows $m_c \sim 0.04 m_e$, the mass drops to $m_c \sim 0.007 m_e$ near the Dirac point ($n_s \sim 2 \times 10^{11} \text{ cm}^{-2}$), where m_e is the mass of the free electron. Overall, the observed gate voltage-dependent effective mass can be fitted to a fictitious 'relativistic' mass: $m_c = E_F/v_F^2 = \sqrt{(\pi \hbar^2 n_s)/v_F^2}$ by using v_F as the only fitting parameter (Fig. 3a, right inset). In accordance with the Berry's phase argument, this procedure extrapolates to a vanishing mass at the Dirac point.

Thus, we have experimentally discovered an unusual QHE in high-quality graphene samples. In contrast with conventional 2D systems, in graphene the observed quantization condition is described by half-integer rather than integer values. The measured phase shift in magneto-oscillation can be attributed to the peculiar topology of the graphene band structure with a linear dispersion relation and vanishing mass near the Dirac point, which can be described in terms of fictitious 'relativistic' carriers. The unique behaviour of electrons in this newly discovered 2 + 1-dimensional quantum electrodynamics system not only opens up many interesting questions in mesoscopic transport in electronic systems with non-zero Berry's phase but may also provide the basis for new applications in carbon-based electric and magnetic field-effect devices, such as ballistic metallic/semiconducting graphene ribbon devices⁹ and electric field effective spin transport devices using a spin-polarized edge state²⁰.

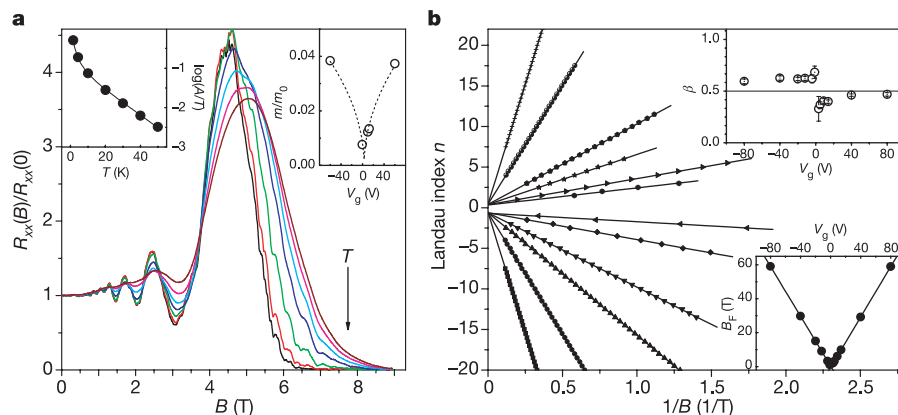


Figure 3 | Temperature dependence and gate-voltage dependence of the SdH oscillations in graphene. **a**, Temperature dependence of the SdH oscillations at $V_g = -2.5$ V. Each curve represents $R_{xx}(B)$ normalized to $R_{xx}(0)$ at a fixed temperature. The curves are in order of decreasing temperature, starting from the top, as indicated by the vertical arrow. The corresponding temperatures are shown in the left inset, which represents the SdH oscillation amplitude A divided by temperature measured at a fixed magnetic field. The standard SdH fit yields the effective mass. The right inset is a plot of the effective mass obtained at different gate voltages. The broken

line is a fit to the single-parameter model described in the text, which yields $v_F = 1.1 \times 10^6 \text{ m s}^{-1}$, in reasonable agreement with published values.

b, A fan diagram for SdH oscillations at different gate voltages. The location of $1/B$ for the n th minimum (maximum) of R_{xx} , counting from $B = B_F$, plotted against n ($n + 1/2$). The lines correspond to a linear fit, in which the slope (lower inset) indicates B_F and the n -axis intercept (upper inset) provides a direct probe of Berry's phase in the magneto-oscillation in graphene. The error bars indicate the standard deviation of fitting errors.

Received 18 July; accepted 12 September 2005.

1. Semenoff, G. W. Condensed-matter simulation of a three-dimensional anomaly. *Phys. Rev. Lett.* **53**, 2449–2452 (1984).
2. Haldane, F. D. M. Model for a quantum hall effect without Landau levels: condensed-matter realization of the “parity anomaly”. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
3. Zheng, Y. & Ando, T. Hall conductivity of a two-dimensional graphite system. *Phys. Rev. B* **65**, 245420 (2002).
4. Gusynin, V. P. & Sharapov, S. G. Unconventional integer quantum Hall effect in graphene. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0506575>) (2005).
5. Peres, N. M. R., Guinea, F. & Neto, A. H. C. Electronic properties of two-dimensional carbon. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0506709>) (2005).
6. Ando, T., Nakaishi, T. & Saito, R. Berry’s phase and absence of back scattering in carbon nanotubes. *J. Phys. Soc. Jpn.* **67**, 2857–2862 (1998).
7. Mikitik, G. P. & Sharlai, Y. V. Manifestation of Berry’s phase in metal physics. *Phys. Rev. Lett.* **82**, 2147–2150 (1999).
8. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
9. Berger, C. *et al.* Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics. *J. Phys. Chem. B* **108**, 19912–19916 (2004).
10. Zhang, Y., Small, J. P., Pontius, W. V. & Kim, P. Fabrication and electric-field-dependent transport measurements of mesoscopic graphite devices. *Appl. Phys. Lett.* **86**, 073104 (2005).
11. Zhang, Y., Small, J. P., Amori, E. S. & Kim, P. Electric field modulation of galvanomagnetic properties of mesoscopic graphite. *Phys. Rev. Lett.* **94**, 176803 (2005).
12. Bunch, J. S., Yaish, Y., Brink, M., Bolotin, K. & McEuen, P. L. Coulomb oscillations and Hall effect in quasi-2D graphite quantum dots. *Nano Lett.* **5**, 287–290 (2005).
13. Roukes, M. L., Scherer, A. & Van der Gaag, B. P. Are transport anomalies in ‘electron waveguides’ classical? *Phys. Rev. Lett.* **64**, 1154–1157 (1990).
14. Fang, Z. *et al.* The anomalous Hall effect and magnetic monopoles in momentum space. *Science* **302**, 92–95 (2003).
15. McEuen, P. L., Bockrath, M., Cobden, D. H., Yoon, Y. & Louie, S. G. Disorder, pseudospins, and backscattering in carbon nanotubes. *Phys. Rev. Lett.* **83**, 5098–5101 (1999).
16. Sharapov, S. G., Gusynin, V. P. & Beck, H. Magnetic oscillations in planar systems with the Dirac-like spectrum of quasiparticle excitations. *Phys. Rev. B* **69**, 075104 (2004).
17. Luk’yanchuk, I. A. & Kopelevich, Y. Phase analysis of quantum oscillations in graphite. *Phys. Rev. Lett.* **93**, 166402 (2004).
18. Morozov, S. V. *et al.* Two dimensional electron and hole gases at the surface of graphite. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0505319>) (2005).
19. Shoenberg, D. *Magnetic Oscillation in Metals* (Cambridge Univ. Press, Cambridge, 1984).
20. Kane, C. L & Mele, E. J. Quantum spin Hall effect in graphene. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0411737>) (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Aleiner, A. Millis, T. F. Heinz, A. Mitra, J. Small and A. Geim for discussions. This research was supported by the NSF Nanoscale Science and Engineering Center at Columbia University, New York State Office of Science (NYSTAR) and the Department of Energy (DOE).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.K. (pkim@phys.columbia.edu).

LETTERS

A record of Permian subaqueous vent activity in southeastern Brazil

Jorge Kazuo Yamamoto¹, Thomas Rich Fairchild¹, Paulo Cesar Boggiani¹, Tarcísio Jose Montanheiro², Carlos César de Araújo³, Pedro Kunihiro Kiyohara⁴, Sergio Luis Fabris de Matos¹ & Paulo César Soares⁵

The remarkable occurrence of more than 4,500 conical siliceous mounds in an area of less than 1.5 square kilometres has been reported in the Paraná basin, near Anhembi, São Paulo, in south-eastern Brazil¹. These structures, which are up to two metres high, are thought to have been formed at the margin of a very shallow, broad but waning internal sea¹, and it was originally suggested that they are stromatolites². Yet their restricted occurrence, unusual abundance and nearly pure siliceous composition have never been satisfactorily explained by this hypothesis. Here we report field and laboratory observations on their shape, construction, composition and mineralogy. On the basis of our data we suggest that the conical mounds are the result of subaqueous Late Permian vent activity in southwestern Gondwana. The present siliceous cone field differs considerably from other Palaeozoic siliceous hot spring deposits, such as those at Rhynie, Scotland³, and the Drummond basin, Australia⁴, and therefore represents an unusual occurrence of vent activity.

The mounds are associated with subhorizontal, thinly bedded siltstones, fine sandstones, and minor limestones of the Teresina Formation (Passa Dois Group, Permian)^{1,5}. Straight-crested ripples, interference ripples, desiccation cracks and mollusc shell concentrations within the section point to a very shallow depositional setting¹. Freshwater charophyte oogonia⁶ are preserved as moulds in siltstone a few metres below the siliceous domes.

Most mounds are steep-sided, rounded cones, 0.2 to 4 m (averaging 0.85 m) across and up to 1.9 m high (Fig. 1), exhibiting an



Figure 1 | Siliceous cones of the Teresina Formation (Late Permian period, Paraná basin, Brazil) exposed in pasture.

originally smooth outer surface with low (≤ 2 cm), broad bumps. They occur individually or in clusters of up to four closely linked cones in such abundance as to sustain local relief. The rock is a variegated light-grey chert ($>98\%$ SiO_2), with abundant vesicles and open spaces partially filled by botryoidal (reminiscent of clustered grapes) chert and radiating, coarse quartz crystals. Petrographically, microcrystalline quartz with salt-and-pepper texture predominates, and the original voids are lined or occluded by subordinate radial-fibrous (optically length-fast) chalcedony and megaquartz (see Supplementary Information).

Weathering, erosion and partial collapse during exhumation have 'decapitated' many cones or, in extreme cases, reduced them to low-walled rings (Fig. 2). The cones consist of a relatively massive outer wall, up to 50 cm thick in the larger structures, and an extremely vesicular or hollow central core extending the entire height of the structure (Fig. 3). It seems that only the thicker-walled forms, of which there are a great many, have retained their apices. Some walls exhibit a faintly defined outermost ring about 10 cm thick or, much more rarely, a few irregular centimetre-thick rings. No finer lamination is evident.

Almost all cones possess some sort of central, apical opening. The vault-like, hollow half-cone in Fig. 3 clearly shows a porous vesicular plug at its apex, but other cones may have originally remained open.



Figure 2 | Top view of siliceous cone being exhumed by modern erosion. Note onlap of terrigenous strata onto the flanks of the cone to the left of the scale (scale, 5 cm).

¹Institute of Geosciences, University of São Paulo, Rua do Lago, 562, CEP 05508-080, São Paulo, Brazil. ²Geological Institute, State Secretary of Environment, Av. Miguel Stefano, 3900, CEP 04301-903, São Paulo, Brazil. ³PETROBRAS SA, Av. República do Chile, 65, CEP 20031-912, Rio de Janeiro, Brazil. ⁴Institute of Physics, University of São Paulo, Travessa R, 187, CEP 05508-900, São Paulo, Brazil. ⁵Department of Geology, Federal University of Paraná, CP 19027, CEP 81531-990, Curitiba, Brazil.



Figure 3 | Half-cone revealing hollow core and distinct vesicular apical plug (just above 10 cm scale bar).

In several uprooted cones, massive or domed, laminated siltstone fills the central core.

A thin siliceous bed, similar to the chert comprising the cones, extends conformably outward from the base of these structures. As observed in outcrop and in several toppled specimens, sedimentary rocks directly beneath some cones have been intensely silicified to a depth of half a metre. Although silica has replaced a carbonate bed containing bivalve shells, intraclasts, ooids and nodular oncoids beneath at least one large overturned cone, there is no petrographic evidence within the cones of corroded carbonate crystals, carbonate crystal fabrics, or irregularly silicified patches that might attest to silica permineralization or replacement of former carbonate.

The shape of the structures is reminiscent of conical stromatolites, modern termite mounds, incipient salt domes, and hydrothermal vent structures from deep-sea, shallow-water and subaerial (geyser) settings. Why are they not stromatolites, as suggested originally², perhaps similar to the coniform stromatolite *Conophyton*, common in the Proterozoic? First, Phanerozoic coniform stromatolites are very rare⁷, and *Conophyton* itself developed primarily at water depths below the fair-weather wave base⁸, deeper than the depositional setting inferred for the siliceous cones near Anhembi. Second, the concentric wall structure observed in some cones is much too rare, gross and irregular to be stromatolitic layering. Third, unlike silicified stromatolites⁹, the cones exhibit no petrographic or scanning electron microscope (SEM) evidence of original calcareous composition or palimpsest microbial lamination. Poorly preserved, grainy-walled filamentous microfossils in tiny tufts are locally abundant in the thin chert bed around the base of the siliceous cones, but no microbial mat-formers or passive inhabitants are evident within the cones themselves.

Insect mounds and salt domes are improbable analogues because the structures in question are symsedimentary features of a shallow-water succession that must have been hardened prior to burial, as overlying sedimentary rocks show compaction deformation and, in at least one case, underlying sediment apparently was forced upward by compaction into the hollow base of the cone. Additionally, these structures are not associated with palaeosols, evaporite minerals or moulds, or any sort of diapiric brecciation or displacive evaporitic crystal growth.

Given the intracratonic nature of the Paraná basin, it is unlikely that the bodies represent silica-replaced chimneys of black or white smokers¹⁰, like those formed in deep marine settings at active plate margins. Indeed, petrography, X-ray diffraction, SEM and chemical analyses of the silica failed to detect smoker wall minerals, such as

Cu–Fe sulphide, anhydrite or barite^{11,12}, or their alteration products (such as limonite).

Exclusion of the deep-sea smoker hypothesis, however, does not eliminate an origin by shallower or even subaerial (geyser) siliceous hydrothermal venting. For example, the cones consist of very pure silica, some of which reveals a relict opaline texture in SEM, suggestive of primary precipitation. A thin, locally microfossiliferous sinter-like chert bed extends outward from the base of many cones. Beds immediately beneath some structures are thoroughly silicified. Terrigenous strata abut against but do not occur within the cones. The mounds are very similar, with a steep, generally massive outer wall, a single vent-like apical opening or porous vesicular plug and a hollow to highly vesicular central core (Fig. 3). Also, they are markedly restricted both geographically and stratigraphically to a single locality within the Paraná sedimentary basin.

From these observations and the lack of evidence of incremental growth or carbonate replacement, we conclude that primary precipitation from relatively continuous, point-sourced discharges of silica-rich waters rapidly formed rigid, erosion-resistant cones during a local, short-lived hydrothermal event. The final stages of this activity eventually plugged the cones and locally silicified underlying sediments. These cones, however, exhibit neither the complex morphology nor the diverse array of facies typical of modern subaerial geyser systems^{13,14} but are similar in composition, their steep-sided, high relief, their narrow vent diameter, internal structure and abundance to inactive siliceous spires up to 7 m high and 2 m wide, thousands of which are present at water depths of 15 m in Yellowstone Lake (Wyoming, USA)¹⁴. We interpret the Anhembi cone field, therefore, as forming by subaqueous hydrothermal venting within the shallow inland sea of the Paraná basin.

Such activity, however, requires water, silica, a plumbing system and a heat source¹³. Water and a source of silica were certainly available in the fine sedimentary rocks of the Teresina Formation. Poor stratigraphic exposure hinders identification of possible conduits in underlying strata, but perhaps the zone of silicification beneath some of the cones is related to the plumbing system. Given the Paraná basin's intracratonic setting and lack of Permian volcanism, the major problem we face is identifying the heat source for the vent activity. Ash falls are registered in the Rio Bonito¹⁵ and Irati formations¹⁶, but they originated from volcanoes nearly 2,000 km from the study area at the active margin of Southwest Gondwana in Argentina¹⁶. Despite this, the interpretation that is best supported by the available data remains the shallow-water siliceous vent model.

Received 12 April; accepted 24 August 2005.

1. Matos, S. L. F. O Contato Entre o Grupo Passa Dois e a Formação Pirambóia na Borda Leste da Bacia do Paraná no Estado de São Paulo. 1–110, Master's thesis, Instituto de Geociências, Univ. São Paulo (1995).
2. Soares, P. C. Estruturas estromatolíticas do Permiano no estado de São Paulo. Res. Comun. Sess. Técnic. 1, 249 (Congresso Brasileiro de Geologia 26, Sociedade Brasileira de Geologia, Belém, 1972).
3. Rice, C. M., Trewin, N. H. & Anderson, L. I. Geological setting of the Early Devonian Rhynie chert, Aberdeenshire, Scotland: an early terrestrial hot spring system. *J. Geol. Soc. Lond.* 159, 203–214 (2002).
4. Cunneen, R. & Sillitoe, R. H. Paleozoic hot spring sinter in the Drummond Basin, Queensland, Australia. *Econ. Geol.* 84, 135–142 (1989).
5. Souza, P. A. & Marques-Toigo, M. An overview on the palynostratigraphy of the Upper Paleozoic strata of the Brazilian Paraná Basin. *Rev. Mus. Argentino Cienc. Nat. (Nueva Serie)* 5, 205–214 (2003).
6. Ragonha, E. W. & Soares, P. C. Ocorrências de carófitas fósseis na Formação Estrada Nova em Anhembi – SP. *Anais* 2, 271–275 (Congresso Brasileiro de Geologia 28, Sociedade Brasileira de Geologia, Porto Alegre, 1974).
7. Walter, M. R., Bauld, J. & Brock, T. D. in *Stromatolites* (ed. Walter, M. R.) 273–310 (Elsevier, Amsterdam, 1976).
8. Donaldson, J. A. in *Stromatolites* (ed. Walter, M. R.) 523–534 (Elsevier, Amsterdam, 1976).
9. Hofmann, H. J. Stratiform Precambrian stromatolites, Belcher Islands, Canada: Relations between silicified microfossils and microstructure. *Am. J. Sci.* 275, 1121–1132 (1975).
10. Tivey, M. K. The formation of mineral deposits at mid-ocean ridges. *Oceanus* 41(2), 22–26 (1998).

11. Haymon, R. M. Growth history of hydrothermal black smoker chimneys. *Nature* **301**, 695–698 (1983).
12. Thurman, H. V. & Burton, E. A. *Introductory Oceanography* 1–554 (Prentice Hall, Englewood Cliffs, New Jersey, 2001).
13. Bryan, T. S. *The Geysers of Yellowstone* 1–462 (Univ. Press of Colorado, Niwot, 1995).
14. Guidry, S. A. & Chafetz, H. S. Anatomy of siliceous hot springs: examples from Yellowstone National Park, Wyoming, USA. *Sedim. Geol.* **157**, 71–106 (2003).
15. Matos, S. L. F., Yamamoto, J. K., Riccomini, C., Hachiro, J. & Tassinari, C. C. Absolute dating of Permian ash-fall in the Rio Bonito Formation, Paraná Basin, Brazil. *Gondwana Res.* **4**, 421–426 (2001).
16. Coutinho, J. M. V., Hachiro, J., Coimbra, A. M. & Santos, P. R. Ash-fall-derived vitroclastic tuffaceous sediments in the Permian of the Paraná Basin and their provenance. *Proc. Int. Gondwana Symp.* **7**, 147–160 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and J. V. M. Coutinho.

Author Contributions J.K.Y. was the project leader, involved in all phases of the research; T.R.F. and P.C.B. did field and laboratory work, discussion and writing; T.J.M. and C.C.A. carried out field work; P.K.K. did laboratory work; S.L.F.M. and P.C.S. did field work and discussion.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.K.Y. (jkyamamo@usp.br).

LETTERS

Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model

Holger Braun¹, Marcus Christl¹, Stefan Rahmstorf², Andrey Ganopolski², Augusto Mangini¹, Claudia Kubatzki³, Kurt Roth⁴ & Bernd Kromer¹

Many palaeoclimate records from the North Atlantic region show a pattern of rapid climate oscillations, the so-called Dansgaard–Oeschger events, with a quasi-periodicity of $\sim 1,470$ years for the late glacial period^{1–6}. Various hypotheses have been suggested to explain these rapid temperature shifts, including internal oscillations in the climate system and external forcing, possibly from the Sun⁷. But whereas pronounced solar cycles of ~ 87 and ~ 210 years are well known^{8–12}, a $\sim 1,470$ -year solar cycle has not been detected⁸. Here we show that an intermediate-complexity climate model with glacial climate conditions simulates rapid climate shifts similar to the Dansgaard–Oeschger events with a spacing of 1,470 years when forced by periodic freshwater input into the North Atlantic Ocean in cycles of ~ 87 and ~ 210 years. We attribute the robust 1,470-year response time to the superposition of the two shorter cycles, together with strongly nonlinear dynamics and the long characteristic timescale of the thermohaline circulation. For Holocene conditions, similar events do not occur. We conclude that the glacial 1,470-year climate cycles could have been triggered by solar forcing despite the absence of a 1,470-year solar cycle.

The onset of successive Dansgaard–Oeschger (DO) events, as documented in Greenland ice-cores^{1,2} for example, is typically spaced by $\sim 1,470$ years or integer multiples thereof^{13,14}. Because deviations from this cyclicity are small, often less than 100–200 years¹⁵, external forcing (solar or orbital) was suggested to trigger DO events^{6,15,16}. However, neither orbital nor solar forcing shows a 1,470-year frequency. Spectral analysis performed on records of cosmogenic nuclides^{8–11}, which are commonly used as proxies for solar variability¹², indicates the possible existence of pronounced and stable^{10,11} centennial-scale solar cycles (the DeVries–Suess and Gleissberg cycles with periods near 210 and 87 years^{10,11}) but does not reveal a 1,470-year cycle⁸. However, the DeVries and Gleissberg cycles are close to prime factors of 1,470 years ($1,470/7 = 210$; $1,470/17 \approx 86.5$). The superposition of two such frequencies could result in variability that repeats with a 1,470-year period.

Here we propose that these two solar frequencies could have synchronized the glacial 1,470-year climate cycle. Support for the idea that a multi-century climate cycle might be linked with century-scale solar variability comes from Holocene data: a multi-centennial drift-ice cycle in the North Atlantic was reported¹⁷ to coincide with “rapid (100- to 200-year), conspicuously large-amplitude variations” in the production rates of the cosmogenic isotopes ¹⁴C and ¹⁰Be. To test our hypothesis, we force the coupled climate system model CLIMBER-2 (version 3) with the two solar frequencies. Earlier simulations with this model showed that, when forced by periodic and/or stochastic variations in the freshwater flux into the northern Atlantic, abrupt glacial warming events are triggered that reproduce

many features of the observed DO events, including the characteristic time evolution and spatial pattern of these events and also the phase relation of the Antarctic response^{18–20}. In the model, the events represent rapid transitions between a stadial (‘cold’) and an interstadial (‘warm’) mode of the North Atlantic thermohaline circulation, triggered by a threshold process.

Following these simulations, we here force the model with variations in freshwater input consisting of two sinusoidal components with periods $T_1 = 1,470/7$ ($= 210$) years and $T_2 = 1,470/17$ (≈ 86.5) years, representing the DeVries and Gleissberg solar cycles. Our focus is the frequency in the glacial climate response (which needs to be studied with a model allowing sufficiently long

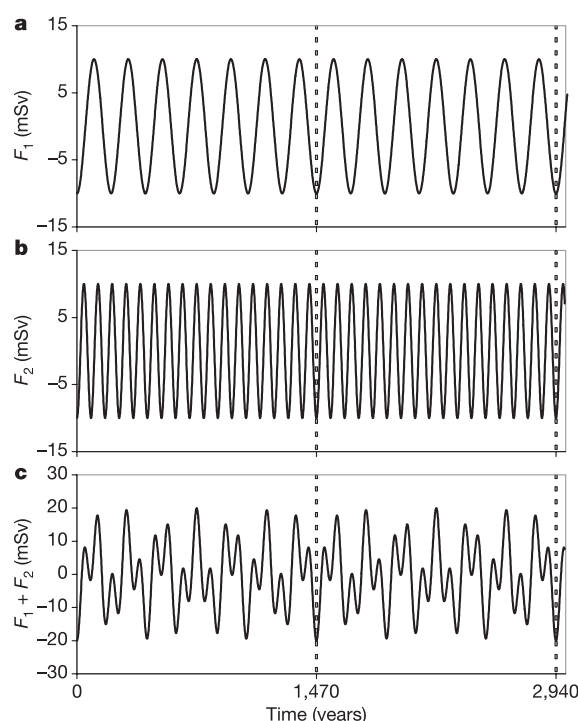


Figure 1 | Applied freshwater forcing in the simulation. **a**, **b**, To represent the DeVries and Gleissberg solar cycles we consider two sinusoidal components: DeVries component F_1 with period $T_1 = 210$ years (**a**) and Gleissberg component F_2 with period $T_2 \approx 86.5$ years (**b**). **c**, Total forcing ($F_1 + F_2$). The dashed lines indicate the period of 1,470 years in the forcing. In the figure, the amplitudes A_1 and A_2 are chosen to be 10 mSv and the phases φ_1 and φ_2 to be 0. The additional offset K is not shown here.

¹Heidelberg Academy of Sciences, c/o Institute of Environmental Physics, University of Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany. ²Potsdam Institute for Climate Impact Research, PO Box 601203, 14412 Potsdam, Germany. ³Alfred Wegener Institute for Polar and Marine Research, Bussestrasse 24, 27570 Bremerhaven, Germany. ⁴Institute of Environmental Physics, University of Heidelberg, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany.

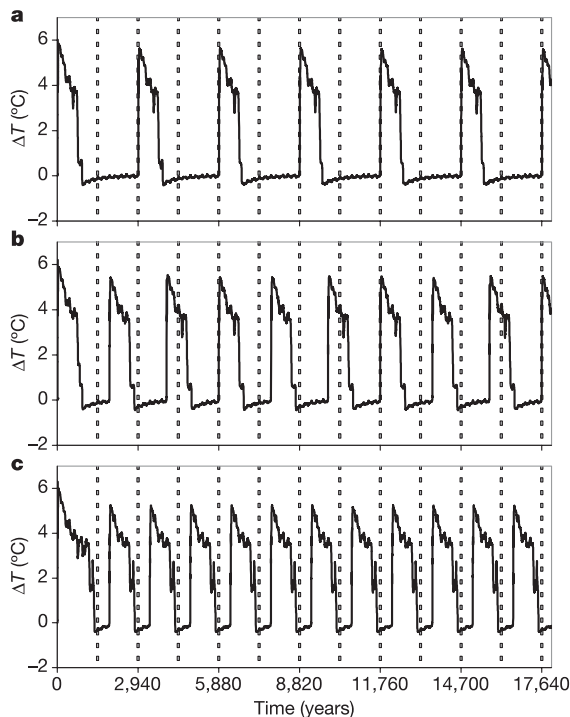


Figure 2 | Simulated changes ΔT in Greenland surface air temperature. The amplitudes A_1 and A_2 of the two forcing cycles are chosen to be 10 mSv, with different values for the offset K . **a**, $K = -9$ mSv; **b**, $K = -14$ mSv; **c**, $K = -19$ mSv. The dashed lines indicate the first-order minima in the forcing (that is, the maxima in the salinity flux), which show a period of 1,470 years (see Fig. 1).

simulations, and correspondingly less resolution and detail). Our study is not aimed at suggesting a certain mechanism for solar influence on freshwater fluxes; this should be studied with more detailed and higher-resolution models. A possible mechanism that has recently been proposed is that solar irradiance variations over the solar periods could generate cyclic salinity anomalies in the North Atlantic by their effect on the thermohaline circulation (THC)²¹. Our simplified approach implies that we assume the known solar frequencies to be present in the hydrological cycle, as suggested

previously^{22,23}. A more direct implementation of solar forcing in the model (such as one based on records of cosmogenic ¹⁰Be) is hindered by several difficulties: for example, a reliable reconstruction of solar activity over tens of thousands of years would be required, with a resolution and dating precision necessary to show the ~ 87 -year Gleissberg cycle realistically. This is so far unachievable during the Last Glacial. Furthermore, when transforming solar forcing into variations of freshwater fluxes, a chain of complex and uncertain processes is involved (for example, atmospheric chemistry^{24–26} and the dynamics of continental ice sheets²⁷ and sea ice²⁸) that cannot explicitly be resolved in climate models designed for millennial-scale simulations.

In our simulations, we consider the Last Glacial Maximum (LGM) as the underlying climate state. The total freshwater forcing F (Fig. 1) reads as follows:

$$F(t) = -A_1 \cos(\omega_1 t + \varphi_1) - A_2 \cos(\omega_2 t + \varphi_2) + K \quad (1)$$

with amplitudes A_1 and A_2 , frequencies $\omega_1 = 2\pi/T_1$ and $\omega_2 = 2\pi/T_2$, and phases φ_1 and φ_2 . The constant offset K represents changes in the background climate compared with the LGM. Following earlier simulations¹⁸, the freshwater perturbation is added to the North Atlantic in the latitude belt 50 – 70° N. This area is particularly sensitive to solar forcing: On one hand, variations in the atmospheric heat fluxes can trigger shifts of the sea-ice margin that—because of the insulating effect of sea-ice²⁸—result in evaporation anomalies. On the other hand, solar forcing can cause meltwater anomalies by its effect on the mass balance of the ice sheets surrounding this region. The applied amplitudes A_1 and A_2 are very small: in the range of 10 mSv (1 Sv = 1 Sverdrup = $10^6 \text{ m}^3 \text{ s}^{-1}$); that is, about 5 cm yr^{-1} in the surface freshwater flux into that area. Although the total forcing does not explicitly have a spectral component of 1,470 years, it repeats with this period because of the combined effect of the applied cycles.

In the response of the model, three different regimes exist: a ‘cold regime’ in which the THC persists in the stadial mode, a ‘warm regime’ in which the interstadial mode is stable, and a ‘Dansgaard–Oeschger regime’ in which cyclic transitions between both modes are excited. These transitions result in abrupt warm events in the model box containing Greenland, similar to DO events (Fig. 2). Figure 3 illustrates the model response for various amplitudes A ($A = A_1 = A_2$) and offsets K . Events with a spacing of 1,470 years are found within a continuous forcing-parameter range. This 1,470-year timescale in the model response is robust when the phases, the

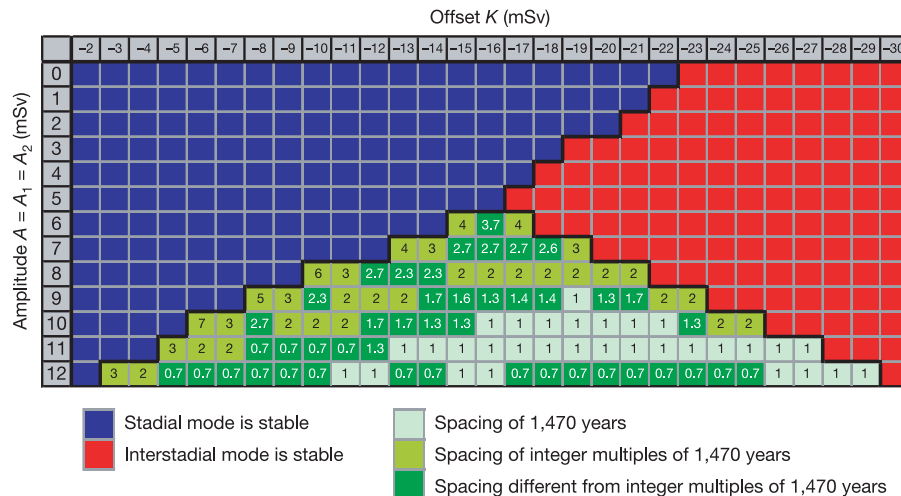


Figure 3 | Schematic response of the model for different amplitudes A (with $A = A_1 = A_2$) and offsets K in the forcing. Upper left area (blue): ‘cold’ regime (stadial mode is stable). Upper right area (red): ‘warm’ regime (interstadial mode is stable). Lower area (green): ‘Dansgaard–Oeschger’

regime (none of the modes is stable). The numbers in this regime indicate the most frequently occurring temporal spacing between successive events (in multiples of 1,470 years) for each combination of A and K .

amplitudes and the periods of the two forcing cycles are changed over some range (Supplementary Information). White noise, when added to the periodic freshwater forcing, acts as an amplifier but does not affect the robustness of the 1,470-year model timescale (Supplementary Information). In the presence of noise, different DO events look different, unlike the identical events occurring for completely regular forcing (Fig. 2). When instead of two sinusoidal cycles a more realistic forcing is applied with spectral properties close to those observed in solar proxies^{8,10,29}, a regular 1,470-year model response can still be obtained (Fig. 4 and Supplementary Information).

The stability of the simulated 1,470-year climate cycle is a plausible consequence of two well-known properties of the THC: its long characteristic timescale, and the high degree of nonlinearity (that is, the threshold character) inherent in the transitions between the two modes of the THC. A very simple conceptual model that only incorporates these two properties is able to mimic key features of CLIMBER-2, for example the existence of three different regimes in the model response, the frequency conversion between forcing and response (that is, the excitation of millennial-scale spectral components in the model response that do not exist in the forcing) and the amplitude dependence of the period in the model response. The general idea of this model is that DO events represent highly nonlinear switches between two different climate states corresponding to the stadial and interstadial modes of the glacial THC. A detailed description of the conceptual model, as well as a discussion of its

response to the applied forcing, is given in the Supplementary Information.

A key aspect of our concept is that the events are triggered by very sharp peaks in the forcing, with duration of only decades. This can explain the puzzling regularity in the timing of DO events in the Greenland data, which were reported to deviate by no more than ~10% from exact multiples of 1,470 years¹⁵. If the events were triggered by broad peaks (for example by some sinusoidal-like forcing cycle of a 1,470-year period), much larger deviations from exact timing would be expected in the 'noisy' climate system. Hence, the exact timing of the more recent DO events (ages between 10,000 and 50,000 years) is a strong observational indication that they are synchronized by shorter cycles.

We have demonstrated that a coupled ocean–atmosphere climate model can reproduce DO events with a robust spacing of 1,470 years when forced by the superposition of two freshwater cycles with much shorter periods near 87 and 210 years. A frequency of 1,470 years is therefore not found in the forcing; it is found only in the model response. We illustrated that this frequency conversion between forcing and response is a plausible consequence of highly nonlinear dynamics inherent in the simulated DO events. Our results indicate that the observed 1,470-year climate cycle could have originated from solar variability despite the lack of a 1,470-year spectral contribution in records of solar activity. Moreover, the 1,470-year climate response in the simulation is restricted to glacial climate and cannot be excited for substantially different (such as Holocene) boundary conditions; for these, the model response shows the frequencies of the applied forcing (~86.5 and 210 years), as also documented in various climate archives^{22,23,30}. Thus, our mechanism for the glacial ~1,470-year climate cycle is also consistent with the lack of a clear and pronounced 1,470-year cycle in Holocene climate archives.

Received 26 May; accepted 10 August 2005.

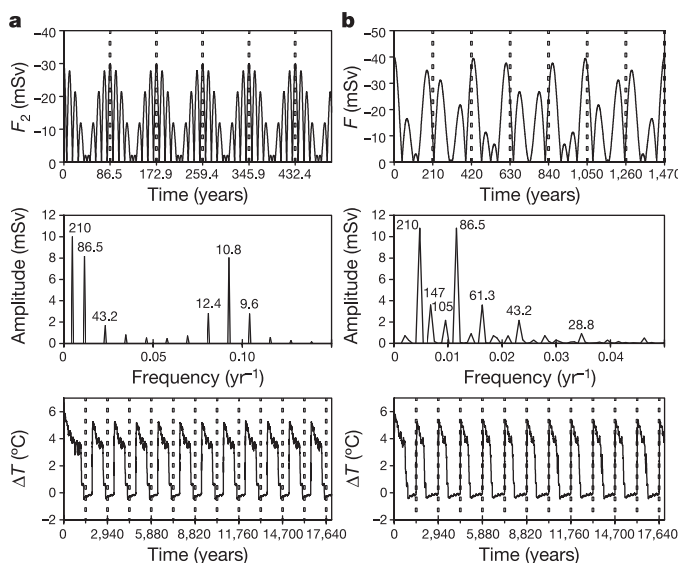


Figure 4 | Model response for non-sinusoidal forcing profiles. a, Roughly 86.5-year Gleissberg cycle implemented as amplitude modulation of a ~11-year cycle (representing the Schwabe solar cycle). Top panel: combined Gleissberg/Schwabe forcing component $F_2 = -A_2 |\cos(\omega_2 t/2)| |\cos(\omega_1 t/2)|$. The chosen forcing parameters are $A_2 = 30$ mSv, $\omega_2 = 2\pi/T_2$ ($T_2 = 1,470/17$ years ≈ 86.5 years), $\omega_1 = 2\pi/T_1$ ($T_1 = 1,470/136$ years ≈ 10.8 years). $|\cos(\omega_1 t/2)|$ is used because $|\cos(\pi t/T)|$ has a period of T . The 210-year DeVries component F_1 is considered as an additional cosine cycle, identical to that in Fig. 1a: $F_1 = -A_1 \cos(\omega_1 t)$ with $A_1 = 10$ mSv and $\omega_1 = 2\pi/T_1$ ($T_1 = 1,470/7$ years = 210 years). Note: here, the freshwater axis is inverted compared with that in Fig. 1. The total forcing is the sum of both components plus an offset $K = -5$ mSv. Middle panel: Fourier spectrum of $F_1 + F_2$. The spectrum is discrete because the forcing is periodic. Numbers above peaks are years. Bottom panel: model response. **b**, Gleissberg cycle modulated by the DeVries cycle. Top panel: combined DeVries/Gleissberg forcing component $F = -A_1 |\cos(\omega_1 t/2)| |\cos(\omega_2 t/2)|$. The chosen forcing parameters are $A_1 = 40$ mSv, $\omega_1 = 2\pi/T_1$ ($T_1 = 1,470/7$ years = 210 years), $\omega_2 = 2\pi/T_2$ ($T_2 = 1,470/17$ years ≈ 86.5 years). The total forcing is given by F plus an offset $K = 5$ mSv. Middle panel: Fourier spectrum of F . Bottom panel: model response.

1. Bond, G. *et al.* Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
2. Grootes, P. M., Stuiver, M., White, J. W. C., Johnson, S. & Jouzel, J. Comparison of the oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
3. Sachs, J. P. & Lehman, S. J. Subtropical North Atlantic temperatures 60,000 to 30,000 years ago. *Science* **286**, 756–759 (1999).
4. Spötl, C. & Mangini, A. Stalagmite from the Austrian Alps reveals Dansgaard–Oeschger events during isotope stage 3: Implications for the absolute chronology of Greenland ice cores. *Earth Planet. Sci. Lett.* **203**, 507–518 (2002).
5. Voelker, A. H. L. *et al.* Global distribution of centennial-scale records for marine isotope stage (MIS) 3: a database. *Quat. Sci. Rev.* **21**, 1185–1212 (2002).
6. Schulz, M. On the 1470-year pacing of Dansgaard–Oeschger warm events. *Paleoceanography* **17**, doi:10.1029/2000PA000571 (2002).
7. van Geel, B. *et al.* The role of solar forcing upon climate change. *Quat. Sci. Rev.* **18**, 331–338 (1999).
8. Stuiver, M. & Braziunas, T. F. Sun, ocean, climate and atmospheric $^{14}\text{CO}_2$: an evaluation of causal and spectral relationships. *Holocene* **3**, 289–305 (1993).
9. Beer, J. *et al.* In *The Sun as a Variable Star: Solar and Stellar Irradiance Variations* (eds Pap, J. M., Fröhlich, H., Hudson, S. & Solanki, S. K.) 291–300 (Cambridge Univ. Press, Cambridge, 1994).
10. Peristykh, A. N. & Damon, P. E. Persistence of the Gleissberg 88-year solar cycle over the last ~12,000 years: Evidence from cosmogenic isotopes. *J. Geophys. Res.* **108**, 1003, doi:10.1029/2002JA009390 (2003).
11. Wagner, G. *et al.* Presence of the solar de Vries cycle (~205 years) during the last ice-age. *Geophys. Res. Lett.* **28**, 303–306 (2001).
12. Lean, J. & Rind, D. Climate forcing by changing solar radiation. *J. Clim.* **11**, 3069–3094 (1998).
13. Alley, R. B., Anandakrishnan, S. & Jung, P. Stochastic resonance in the North Atlantic. *Paleoceanography* **16**, 190–198 (2001).
14. Alley, R. B., Anandakrishnan, S., Jung, P. & Clough, A. In *The Oceans and Rapid Climate Change: Past, Present and Future* (eds Seidov, D., Maslin, M. & Haupt, B. J.) 57–68 (AGU Geophys. Monogr. 126, Washington, D.C., 2001).
15. Rahmstorf, S. Timing of abrupt climate change: A precise clock. *Geophys. Res. Lett.* **30**, 1510, doi:10.1029/2003GL017115 (2003).
16. Keeling, C. D. & Whorf, T. P. The 1,800-year oceanic tidal cycle: A possible cause of rapid climate change. *Proc. Natl Acad. Sci. USA* **97**, 3814–3819 (2000).
17. Bond, G. *et al.* Persistent solar influence on North Atlantic climate during the Holocene. *Science* **294**, 2130–2136 (2001).

18. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
19. Ganopolski, A. & Rahmstorf, S. in *The Oceans and Rapid Climate Change: Past, Present and Future* (eds Seidov, D., Maslin, M. & Haupt, B. J.) 261–275 (AGU Geophys. Monogr. 126, Washington, D.C., 2001).
20. Ganopolski, A. & Rahmstorf, S. Abrupt glacial climate changes due to stochastic resonance. *Phys. Rev. Lett.* **88**, 038501 (2002).
21. Weber, S. L., Crowley, T. J. & van der Schrier, G. Solar irradiance forcing of centennial climate variability during the Holocene. *Clim. Dyn.* **22**, 539–553 (2004).
22. Neff, U. *et al.* Strong coherence between solar variability and the monsoon in Oman between 9 and 6 kyr ago. *Nature* **411**, 290–293 (2001).
23. Hodell, D. A., Brenner, M., Curtis, J. H. & Guilderson, T. Solar forcing of drought frequency in the Maya lowlands. *Science* **292**, 1367–1370 (2001).
24. Haigh, J. D. The role of stratospheric ozone in modulating the solar radiative forcing of climate. *Nature* **370**, 544–546 (1994).
25. Shindell, D., Rind, D., Balachandran, N., Lean, J. & Loneragan, P. Solar cycle variability, ozone, and climate. *Science* **284**, 305–308 (1999).
26. Carslaw, K. S., Harrison, R. G. & Kirkby, J. Cosmic rays, clouds and climate. *Science* **298**, 1732–1737 (2002).
27. Schmittner, A., Yoshimori, M. & Weaver, A. J. Instability of glacial climate in a model of the ocean–atmosphere–cryosphere system. *Science* **295**, 1489–1493 (2002).
28. Lohmann, G. & Gerdes, R. Sea ice effects on the sensitivity of the thermohaline circulation in simplified atmosphere–ocean–sea ice models. *J. Clim.* **11**, 2789–2803 (1998).
29. Ogurtsov, M. G., Nagovitsyn, Y. A., Kocharov, G. E. & Jungner, H. Long-period cycles of the Sun's activity recorded in direct solar data and proxies. *Solar Phys.* **211**, 371–394 (2002).
30. Ram, M. & Stolz, M. Possible solar influences on the dust profile of the GISP2 ice core from Central Greenland. *Geophys. Res. Lett.* **26**, 1043–1046 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Latuske for insight into spectral analysis; M. Claussen, R. Calov and E. Bauer for support and helpful comments; and the late G. Bond for his support and for many discussions that helped to improve the quality of this study. C.K. was funded by a subcontract to a project of the Bundesministerium für Bildung und Forschung (BMBF).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.B. (holger.braun@iup.uni-heidelberg.de).

LETTERS

The deterministic nature of earthquake rupture

Erik L. Olson¹ & Richard M. Allen²

Understanding the earthquake rupture process is central to our understanding of fault systems and earthquake hazards. Multiple hypotheses concerning the nature of fault rupture have been proposed but no unifying theory has emerged^{1–12}. The conceptual hypothesis most commonly cited is the cascade model for fault rupture^{1,3,10,13}. In the cascade model, slip initiates on a small fault patch and continues to rupture further across a fault plane as long as the conditions are favourable. Two fundamental implications of this domino-like theory are that small earthquakes begin in the same manner as large earthquakes and that the rupture process is not deterministic—that is, the size of the earthquake cannot be determined until the cessation of rupture. Here we show that the frequency content of radiated seismic energy within the first few seconds of rupture scales with the final magnitude of the event. We infer that the magnitude of an earthquake can therefore be estimated before the rupture is complete. This finding implies that the rupture process is to some degree deterministic and has implications for the physics of the rupture process.

In the cascade model, faults are divided into patches of varying size and shape. When an earthquake initiates on one patch, slip on that patch can lead to slip on the adjacent patches if the rupture energy and the state of stress on these adjacent patches are favourable. An earthquake continues spreading from patch to patch until there is insufficient energy to rupture the next patch, at which point the rupture stops. This model is consistent with the concepts of seismic moment and magnitude. Moment is physically related to both rupture area, A , and average slip, $\bar{D}$, by the relation $M_0 = \mu A \bar{D}$ where M_0 is the seismic moment and μ is the shear modulus. Moment magnitude, M_w , scales with the seismic moment and is therefore similarly related to A and $\bar{D}$ (ref. 14). Given this framework, it is not possible to know the magnitude of an earthquake until the rupture has stopped.

Throughout the last decade, the seismological community has debated whether the first few seconds of the P wave (the first few seconds of radiated energy) provides information about the final magnitude of an earthquake before the rupture is complete^{3,9–11}. Much of the debate has focused on the time-domain characteristics of the P wave. However, evidence for a scaling relation between the frequency content of the first few seconds of the P wave and the final magnitude has also emerged^{15–18}. Using an approach similar to Nakamura¹⁵, Allen and Kanamori¹⁶ measured the predominant period, τ_p , from the first 4 s of the P-wave arrival at multiple seismic stations in southern California. They showed a scaling relation between τ_p and magnitude M for earthquakes with M of 3.0 to 7.3. In earthquakes with $M < 6$ the total duration of rupture is usually less than 4 s, so the entire rupture time history is included within the first 4 s of the P wave. However, for $M > 6$ earthquakes, the existence of a scaling relation between τ_p and M would imply that the magnitude of an earthquake has been defined before the rupture terminates, and that the rupture process is deterministic. Allen and Kanamori¹⁶ used earthquakes from southern California where data for only three earthquakes with $M_w > 6$ are available.

Here we measure τ_p for a much larger number of earthquakes, including events from Japan, Taiwan, California and Alaska. The waveform data have been provided by K-net (operated by the National Research Institute for Earth Science and Disaster Prevention in Japan), by the Taiwan Strong Motion Instrumentation Program of the Central Weather Bureau, by the Southern California Seismic Network (operated by Caltech and the US Geological Survey), and by the University of Alaska Geophysical Institute. A total of 71 earthquakes producing 1,842 waveforms recorded on both broadband velocity sensors and accelerometers within 100 km of the epicentre are used. There are 24 events with $M_w \geq 6.0$, including the M_w 7.6 Chi-Chi earthquake (Taiwan, 1999) with a rupture duration of ~ 30 s, the M_w 7.9 Denali earthquake (Alaska, 2002) with a rupture duration of ~ 70 s, and the M_w 8.3 Tokachi-oki earthquake (Japan, 2003) with a duration of ~ 40 s. A table of events is included in the Supplementary Information.

We calculate τ_p in a recursive fashion from a vertical velocity timeseries to generate τ_p as a function of time, $\tau_p(t)$ (see Methods section). Figure 1 shows the vertical velocity waveform recorded during a M 4.6 earthquake in southern California and the τ_p timeseries derived from it. Figure 2 shows a similar example but for the M_w 8.3 Tokachi-oki earthquake. In this case only acceleration records are available, and they have been recursively integrated in a causal fashion to derive the velocity trace from which $\tau_p(t)$ is derived. We define the parameter $\tau_p^{\max}$ as the maximum $\tau_p(t)$ data point between 0.05 and 4.0 s (see Methods section) after the P-wave trigger as shown in Figs 1 and 2.

When $\tau_p^{\max}$ is plotted against M on a log-linear scale, a scaling relation emerges as shown in Fig. 3a. The $\tau_p^{\max}$ observations from waveforms at individual stations can exhibit large variability for a single earthquake, which is probably due to measurement error, station effects and path effects¹⁹. Figure 3a shows the average $\tau_p^{\max}$ observation for each earthquake using all available data. The best-fit linear relation to the event averages is:

$$\log \tau_p^{\max} = 0.14M - 0.83 \quad (1)$$

and the average absolute deviation is 0.54 magnitude units. The data set has a high linear correlation coefficient of 0.9. For 90% of the earthquakes in this study, $\tau_p^{\max}$ is within two times the average absolute deviation. In determining a single linear relation to the entire data set, we have chosen the simplest possible model. Given the hint of a break in our data around M 5.7, we also determined best-fit relations for $M > 5.7$ and $M < 5.7$ events. In both cases the slopes are positive, but lower than equation (1), and the correlation coefficients are reduced, particularly for $M > 5.7$. Although there is variability in $\tau_p^{\max}$ observations, equivalent to ± 1 magnitude unit, a scaling relation is clear: this implies that information about the final magnitude of an earthquake is available within the first few seconds of its initiation, irrespective of the total rupture duration.

A second parameter, τ_d , is also measured from the τ_p timeseries. τ_d is the delay of the $\tau_p^{\max}$ observation with respect to the P-wave trigger, and is therefore in the range of 0.05 s to 4 s (see Figs 1 and 2).

¹Department of Geology and Geophysics, University of Wisconsin Madison, 1215 W. Dayton Street, Madison, Wisconsin 53706, USA. ²Department of Earth and Planetary Science, University of California Berkeley, 307 McCone Hall, Berkeley, California 94720, USA.

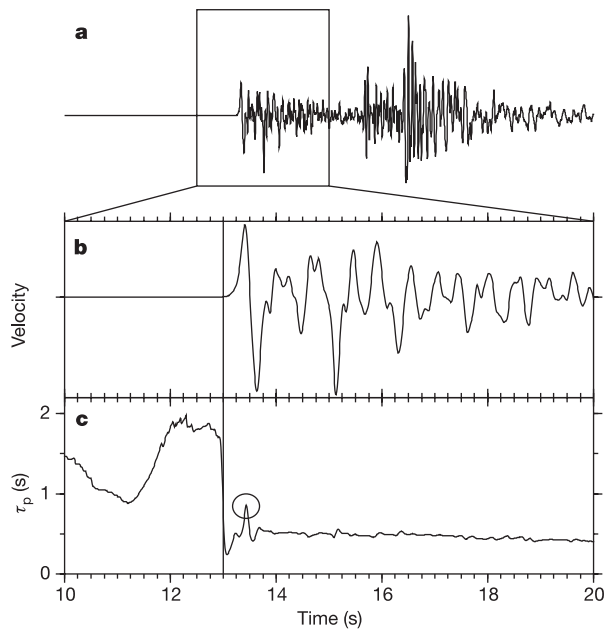


Figure 1 | Example waveform and $\tau_p^{\max}$ calculation for a M 4.6 earthquake in southern California recorded at station GSC, 74 km from the epicentre.

a, The raw vertical component waveform recorded by a broadband velocity sensor. **b**, Ten seconds of the velocity waveform after low-pass filtering at 3 Hz. The P-wave trigger time is shown by the vertical line at 13.01 s. **c**, $\tau_p(t)$ trace calculated in a recursive fashion from the waveform in **b**, showing the change in the frequency content from the pre-trigger noise to the post-trigger P wave. The $\tau_p^{\max}$ observation is circled (equal to 0.86 s in this case); τ_d is the delay of $\tau_p^{\max}$ with respect to the trigger (0.43 s in this case).

Figure 3b plots the event-averaged τ_d observations versus magnitude, showing a general increase in τ_d with magnitude. Also indicated on Fig. 3b is the typical rupture duration as a function of magnitude. The relation shown is only approximate, as the rupture duration for a given magnitude event can vary by a factor of 2 or 3 (see Methods section). Despite the uncertainty in the rupture duration of the specific earthquakes included in this study, it is clear that for earthquakes with $M > 4$ the $\tau_p^{\max}$ observation is made before the rupture has ceased. While up to 4 s of data are used to determine $\tau_p^{\max}$, the average time window of the P wave required to determine $\tau_p^{\max}$ is less than 2 s for almost all earthquakes in our study.

Using published moment rate functions for the $M > 6$ events in our study, we can estimate the amount of moment release at time τ_d . Rupture directivity effects of finite faults mean that the first 2 s of the P wave at a given station does not sample exactly 2 s of the rupture. However, because our τ_d measurement is event-averaged we can use it as an approximate estimate of the rupture duration within which $\tau_p^{\max}$ information is available. Calculating the seismic moment released within τ_d shows an increase with magnitude, which is not surprising given that τ_d also increases with magnitude. But the percentage of moment released is always small (less than 40% for $M > 6$ events) and decreases with magnitude. In the case of the largest earthquakes—Chi-Chi, Tokachi-oki and Denali—the percentage of moment released is less than 2%. The fraction of the total rupture duration at which point $\tau_p^{\max}$ information is available is not constant, and decreases with increasing magnitude.

Although there is a ± 1 magnitude unit scatter in the $\tau_p^{\max}$ data, the observations show that the rupture process is at least partly deterministic: that is, the final magnitude of an event is to some degree controlled by processes within the first few seconds (typically < 2 s) of rupture. The scatter could be due to source processes and/or local site and measurement errors. If the scatter is non-source related, then removal or correction for site and path effects could reduce the

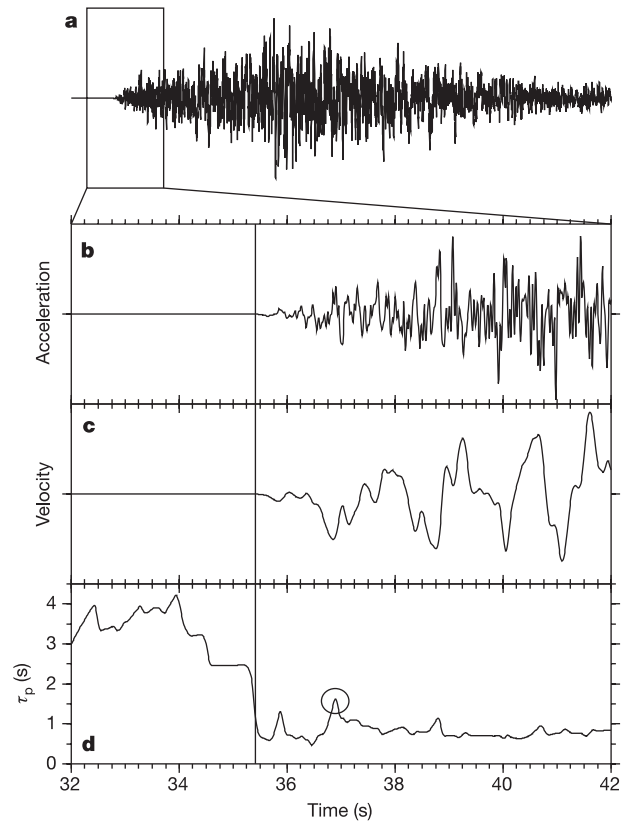


Figure 2 | Example waveform and $\tau_p^{\max}$ calculation for the M_w 8.3 Tokachi-oki earthquake, recorded at station HKD112, 71 km from the epicentre. **a**, The raw vertical component waveform recorded on an accelerometer. **b**, Ten seconds of the raw acceleration waveform. The P-wave trigger is shown by the vertical line at 35.41 s. **c**, Ten seconds of the velocity waveform determined from the acceleration recording using recursive relations only. It has also been low-pass filtered at 3 Hz. **d**, $\tau_p(t)$ trace calculated in a recursive fashion from the waveform in **c**. The $\tau_p^{\max}$ observation is circled ($\tau_p^{\max} = 1.62$ s, $\tau_d = 1.49$ s), it has a longer period and is observed later than the example in Fig. 1c owing to the larger magnitude of the earthquake.

scatter in the data points of Fig. 3a to a single line, implying that the final magnitude of an earthquake is entirely determined within the first few seconds of rupture. Variability in the quality of the $\tau_p^{\max}$ observations at different stations has already been observed across the seismic network in southern California, indicating that site effects do play a role¹⁹. Nevertheless, it seems unlikely that all the scatter is due only to site effects. Instead, source-related processes (including rupture behaviour, stress heterogeneity and other on-fault variability) probably also play a role.

The cascade model of sequential slip on adjacent fault patches must be refined in order to be consistent with our observations. In particular, a cascade in which a small or large earthquake could initiate as slip on any small patch and grow according to the conditions on the surrounding fault plane until rupture arrest is inconsistent with our observed scaling relation. This does not preclude the existence of multiple slip patches within a rupture. For some earthquakes, a sequence of ‘subevents’ (which could be considered as slip on specific patches) has been observed within the τ_d time window and $\tau_p^{\max}$ is a measure of the final subevent—for example, the Landers earthquake²⁰. In other cases, however, $\tau_p^{\max}$ is observed before completion of the first subevent, and before the later, larger subevents are under way—as in the case of the Denali earthquake²¹. Therefore, although rupture may still be considered as slip on a series of fault patches, the final seismic moment is at least partially determined by the patch or patches that rupture within the first few seconds.

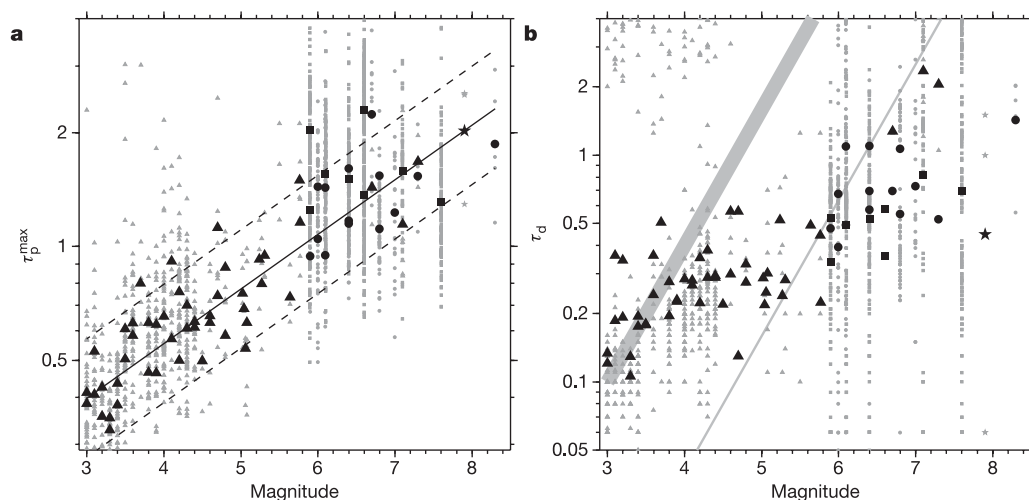


Figure 3 | The relation between $\tau_p^{\max}$, τ_d and magnitude. **a**, The scaling relation between $\tau_p^{\max}$ and magnitude for earthquakes in southern California (triangles), Japan (circles), Taiwan (squares) and Alaska (stars). Observations at individual stations can show a large scatter for a given event (small grey shapes). The event averaged values are also shown (large black shapes). The best-fit to event averages is shown as a solid line. The average absolute deviation of the observations is 0.54 magnitude units; plus and minus two times the average absolute deviation is shown as dashed lines.

b, τ_d plotted against magnitude, showing the general increase in the time required to make the $\tau_p^{\max}$ observation with increasing magnitude. The symbols are the same as in **a**. Grey shapes indicate individual station observations, black is the event average. The thick grey bar shows the approximate rupture duration as a function of magnitude, indicating that the $\tau_p^{\max}$ observation is made before rupture ceases for all $M > 4$ earthquakes in this study. The thin grey line indicates one-tenth of the rupture duration, and is crossed by the trend of the τ_d data.

We propose that the final magnitude of an earthquake is partially controlled by the initiation process within the first few seconds of rupture, and partially by the physical state of the surrounding fault plane. The role played by the initiation process can be understood by considering the energy balance of fault rupture. A rupture can only propagate when the available energy is sufficient to supply the necessary fracture energy^{22,23}. When a propagating fracture encounters a patch with a lower stress-drop, the total energy in the system will begin to decrease. Depending on the size of the patch, it may cause the rupture to terminate. The total rupture energy available increases with the amount of slip, so a large-slip rupture will propagate further across a heterogeneous fault plane. Therefore, if the rupture pulse initiates with large slip, it is more likely to evolve into a large earthquake. This explanation is consistent with the observation that large earthquakes do not nucleate at shallow depths, but instead at greater depths where the frictional strength and stress drop are greater²⁴. A recent study²⁵ also shows that hypocentres are preferentially located within or close to regions of large slip.

The $\tau_p^{\max}$ scaling relation provides new constraints on the physics of the rupture process. Whereas the cascade model suggests that the magnitude of an earthquake is dependent on the state of stress across the fault plane and that the nucleation of all earthquakes is identical, our present observations demonstrate the significance of the rupture initiation process within the first few seconds of an event. Understanding the physics of this process will enable us to predict the magnitude of earthquakes without accurate knowledge of the surrounding state of stress across a fault plane.

METHODS

Waveform processing. The $\tau_p^{\max}$ observations are derived from the vertical component of either broadband velocity seismometers or accelerometers. The acceleration waveforms are converted to velocity, and all data are low-passed at 3 Hz using the recursive relations described in ref. 26. The $\tau_p(t)$ timeseries is then calculated from the low-passed velocity waveform using the following recursive relation:

$$\tau_i^p = 2\pi \sqrt{X_i/D_i}$$

Here $X_i = \alpha X_{i-1} + x_i^2$, $D_i = \alpha D_{i-1} + (dx/dt)_i^2$, x_i is the ground motion recorded at time i and α is a 1 s smoothing constant. For 100 sample per second (s.p.s.) data $\alpha = 0.99$, for 20 s.p.s. data $\alpha = 0.95$. This approach was first described in ref. 15; more information is available in ref. 16.

We define the parameter $\tau_p^{\max}$ as the maximum $\tau_p(t)$ data point between 0.05 s and 4.0 s after the P-wave trigger, as shown in Figs 1 and 2. The time window starts at 0.05 s rather than 0.00 s owing to the recursive nature of the $\tau_p(t)$ calculation. Using the maximum value between 0.0 s and 4.0 s can result in leakage of the frequency content of background noise before the P-wave arrival into the time window after the P wave. We also experimented with time windows ranging in duration from 1 s to 5 s. The upper limit was set to 5 s, as the S-wave arrival, which contaminates the P-wave signal, is 5 s behind the P wave at an epicentral distance of ~ 50 km, half the epicentral distance range of the data. Experimenting with these different time windows, we found that there was no improvement in the relation between $\tau_p^{\max}$ and magnitude when 5 s of data were used rather than 4 s. There was, however, improvement in the relation when 4 s was used rather than a shorter time window.

Estimation of rupture duration. The relation between rupture duration and magnitude shown in Fig. 3b is based on scaling relations between rupture length and magnitude. The rupture length is converted to rupture duration by assuming unilateral rupture and a rupture velocity between 2.4 km s^{-1} and 3.0 km s^{-1} from the estimates of ref. 27. Two approaches to estimating rupture length are applied. First, the empirical scaling relations between magnitude and rupture length developed in ref. 28 are used. These are appropriate for $M > 4.5$ earthquakes. Second, we use the scaling relations between rupture dimension and seismic moment for stress drops in the range 0.1–10 MPa. The rupture durations shown in Fig. 3b are an average of the range of rupture durations obtained using the approaches described above. The actual rupture duration for a given magnitude event can vary by a factor of 2 or 3.

Received 12 April; accepted 30 August 2005.

- Brune, J. N. Implications of earthquake triggering and rupture propagation for earthquake prediction based on premonitory phenomena. *J. Geophys. Res.* **84**, 2195–2198 (1979).
- Fukao, Y. & Furumoto, M. Hierarchy in earthquake size distribution. *Phys. Earth Planet. Inter.* **37**, 149–168 (1985).
- Ellsworth, W. L. & Beroza, G. C. Seismic evidence for an earthquake nucleation phase. *Science* **268**, 851–855 (1995).
- Beroza, G. C. & Ellsworth, W. L. Properties of the seismic nucleation phase. *Tectonophysics* **261**, 209–227 (1996).
- Dodge, D. A., Beroza, G. C. & Ellsworth, W. L. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res.* **101**, 22371–22392 (1996).
- Mori, J. Rupture directivity and slip distribution of the M 4.3 foreshock to the 1992 Joshua Tree earthquake, Southern California. *Bull. Seismol. Soc. Am.* **86**, 805–810 (1996).
- Mori, J. & Kanamori, H. Initial rupture of earthquakes in the 1995 Ridgecrest, California sequence. *Geophys. Res. Lett.* **23**, 2437–2440 (1996).
- Singh, S. K. *et al.* Implications of a composite source model and seismic-wave

- attenuation for the observed simplicity of small earthquakes and reported duration of earthquake initiation phase. *Bull. Seismol. Soc. Am.* **88**, 1171–1181 (1998).
9. Steacy, S. J. & McCloskey, J. What controls an earthquake's size? Results from a heterogeneous cellular automaton. *Geophys. J. Int.* **133**, F11–F14 (1998).
 10. Kilb, D. & Gombert, J. The initial subevent of the 1994 Northridge, California, earthquake: Is earthquake size predictable? *J. Seismol.* **3**, 409–420 (1999).
 11. Sato, T. & Kanamori, H. Beginning of earthquakes modeled with the Griffith's fracture criterion. *Bull. Seismol. Soc. Am.* **89**, 80–93 (1999).
 12. Ohnaka, M. A physical scaling relation between the size of an earthquake and its nucleation zone size. *Pure Appl. Geophys.* **157**, 2259–2282 (2000).
 13. Ellsworth, W. L. & Beroza, G. C. Observation of the seismic nucleation phase in the Ridgecrest, California, earthquake sequence. *Geophys. Res. Lett.* **25**, 401–404 (1998).
 14. Kanamori, H. The energy release in great earthquakes. *J. Geophys. Res.* **82**, 2981–2987 (1977).
 15. Nakamura, Y. in *Proc. 9th World Conf. Earthquake Eng.* VII, 673–678 (1988).
 16. Allen, R. M. & Kanamori, H. The potential for earthquake early warning in southern California. *Science* **300**, 786–789 (2003).
 17. Kanamori, H. The diversity of the physics of earthquakes. *Proc. Jpn. Acad. B* **80**, 297–316 (2004).
 18. Nakamura, Y. in *Proc. 13th World Conf. Earthquake Eng.* Paper No. 908 (2004).
 19. Lockman, A. & Allen, R. M. Single station earthquake characterization for early warning. *Bull. Seismol. Soc. Am.* (in the press).
 20. Abercrombie, R. & Mori, J. Local observations of the onset of a large earthquake: 28 June 1992 Landers, California. *Bull. Seismol. Soc. Am.* **84**, 725–734 (1994).
 21. Eberhart-Phillips, D. *et al.* The 2002 Denali Fault Earthquake, Alaska: A large magnitude, slip-partitioned event. *Science* **300**, 1113–1118 (2003).
 22. Nielsen, S. B. & Olsen, K. B. Constraints on stress and friction from dynamic rupture models of the 1994 Northridge, California, earthquake. *Pure Appl. Geophys.* **157**, 2029–2046 (2000).
 23. Oglesby, D. D. & Day, S. M. Stochastic fault stress: Implications for fault dynamics and ground motion. *Bull. Seismol. Soc. Am.* **92**, 3006–3021 (2002).
 24. Das, S. & Scholz, C. H. Why large earthquakes do not nucleate at shallow depths. *Nature* **305**, 621–623 (1983).
 25. Mai, P. M., Spudich, P. & Boatwright, J. Hypocenter locations in finite-source rupture models. *Bull. Seismol. Soc. Am.* **95**, 965–980 (2005).
 26. Kanamori, H., Maechling, P. & Hauksson, E. Continuous monitoring of ground-motion parameters. *Bull. Seismol. Soc. Am.* **89**, 311–316 (1999).
 27. Somerville, P., Irikura, K., Graves, R. P., Sawada, S. & Wald, D. Characterizing crustal earthquake slip models for the prediction of strong motion. *Seismol. Res. Lett.* **70**, 59–80 (1999).
 28. Wells, D. L. & Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.* **84**, 974–1002 (1994).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank H. Kanamori and S. Nielsen for discussions, and Y.-M. Wu and R. Hansen for making waveform data available for the study. The manuscript was improved by comments from R. Abercrombie and C. Scholz. Funding for this work was provided by the US Geological Survey NEHRP programme, the University of Wisconsin, Madison, and the University of California, Berkeley.
- Author Information** Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M.A. (rallen@berkeley.edu).

LETTERS

Single origin of a pan-Pacific bird group and upstream colonization of Australasia

Christopher E. Filardi^{1*†} & Robert G. Moyle^{1*}

Oceanic islands have long served as natural laboratories for understanding the diversification of life^{1–4}. In particular, the many thousands of islands spanning the tropical Pacific support an unparalleled array of terrestrial communities whose patterns of diversity contributed fundamental insights to the development of classical speciation and biogeographic theory^{4–8}. Much of this work is founded on an assumption derived from traditional taxonomic approaches, namely that faunas on these widely separated archipelagos stem from a simple one-way, downstream flow of colonists from continents to islands^{2,4}. Here we show, with the use of molecular phylogenetic data from one of the original bird families used to justify this assumption, that a diverse array of endemic island genera and species are the product of a single radiation that diversified across all major Pacific archipelagos in a non-stepping-stone fashion, and recently recolonized continental areas. The geographic scope and lineage-specific approach of this study reveal evolutionary patterns long obscured by traditional taxonomic surveys and indicate that widely dispersed archipelagos can be sources of biological diversity.

Explaining the origin of regional biotas relies on an accurate estimation of evolutionary relationships among species. Because most geographic patterns of diversity have been defined by classical taxonomy rather than by modern phylogenetic analysis, biogeographic hypotheses may not reflect hierarchical speciation history. This is especially true of oceanic islands, where isolation and the potential for rapid differentiation can obscure evolutionary affinities, which in turn confounds biogeographic interpretation. Despite the disproportionate influence of insular faunas of the tropical Pacific on biogeographic and speciation theory^{2,6,7}, few systematic analyses exist for major pan-Pacific groups^{9,10}. Instead, the interpretation of Pacific faunas mainly derives from taxonomic surveys that demonstrate an attenuation of species diversity with increased distance from continental areas, leading to the assumption that island communities have generally resulted from a simple one-directional flow of colonists from continents to far-flung archipelagos^{2,4,11}.

This assumption is evident in Mayr's seminal interpretations of Pacific bird diversity^{2,11}, which have influenced general theory explaining the origins of insular biotas. Mayr concluded that there is no Polynesian avifauna, there are only colonists from New Guinea; this conclusion is reflected in current taxonomic treatment of island forms, with many small or monotypic genera isolated on archipelagos. Similarly, a recent synthesis of bird speciation patterns across the islands of northern Melanesia emphasized the pervasive influence of downstream colonization from continental Australasia⁸. The geographic source of Pacific endemics should be directly testable

in a phylogenetic framework. If island species and genera are in fact independently derived from continental sources, insular taxa should be sister to continental forms. Alternatively, evidence of extensive endemic radiation within the Pacific avifauna would challenge long-standing assumptions about the origin of insular diversity. Monarch flycatchers (family Monarchidae) feature prominently in Mayr's work and are particularly suited to testing the origin of Pacific faunas. The family contains seven genera endemic to Pacific islands, as well as genera more widespread throughout the Old World tropics. Furthermore, recent phylogenetic results¹² revealed an endemic island genus embedded within the widespread genus *Monarcha*, indicating that current taxonomy may obscure the extent of Pacific monarch radiations.

Using parsimony and model-based phylogenetic methods, we analysed nuclear and mitochondrial DNA sequence data to reconstruct evolutionary relationships among a broad sampling of continental and island monarchs. Two notable and well-supported results emerged in all analyses. First, all six genera endemic to Pacific archipelagos were nested within the genus *Monarcha* (Fig. 1). Also included in this clade of endemic genera (clade A; Fig. 1) were several *Monarcha* species endemic to Pacific islands. This indicates that a large and diverse array of Pacific flycatchers resulted from a single radiation involving nearly every major Pacific archipelago, rather than the gradual accretion of independent lineages from continental sources. Furthermore, the sequence of speciation events within this island clade does not indicate a stepping-stone mode of island colonization. For example, *Chasiempis* and *Pomarea*, two of the most basal taxa, are endemic to the Hawaiian Islands and eastern Polynesia, respectively, which are two of the most remote regions in the Pacific (Fig. 2). In the second key result, phylogenetic evidence suggested two widespread continental species (*M. frater* and *M. melanopsis*) recently diverged from sister taxa embedded well within this pan-Pacific radiation. This directly contradicts the assumption that these taxa are basal to island forms, which is cited as support for the standard continent-to-island immigration pattern⁸.

Not surprisingly, reconstruction of ancestral areas revealed two geographically distinct radiations, affirming visual patterns evident in the phylogeny (Fig. 1). The earliest speciation events are reconstructed as 'continental', and clade B follows the expected pattern of continental lineages giving rise to isolated island taxa. In contrast, inferred ancestral areas within clade A strongly support a history of monarch diversification centred in the tropical Pacific. All of the earliest speciation events are reconstructed as 'island' and the single continental ancestral area is distal and recent, pointing to the recolonization of continental areas by members of an entirely

¹Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA. †Present address: Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA.

*These authors contributed equally to this work.

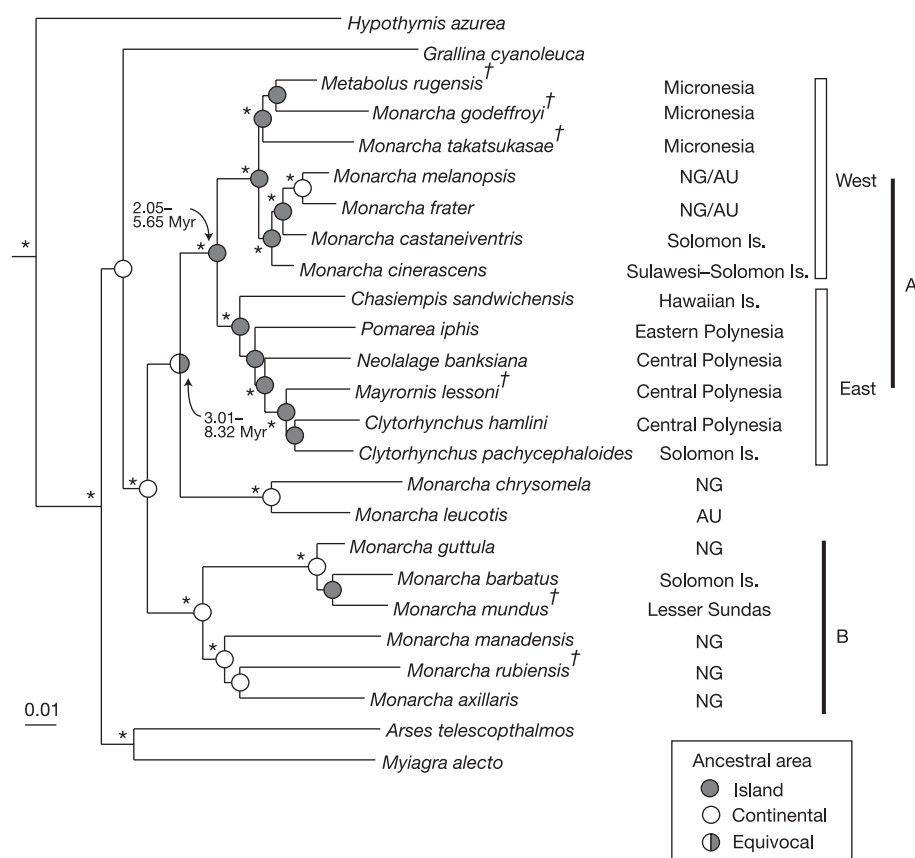


Figure 1 | Maximum-likelihood phylogeny of Pacific monarchs based on ND2 and myoglobin intron-2 data. The tree is rooted with *Machaerirhynchus* and *Rhipidura* (not shown). Daggers indicate taxa for which DNA was extracted from museum study skins and are represented by mitochondrial DNA only. Asterisks indicate nodes receiving more than 0.95 posterior probability and/or more than 70% maximum-likelihood

bootstrap support. General geographic distributions are indicated to the right of taxon names. Ancestral area assignments for nodes are based on maximum-likelihood-based reconstructions (see Supplementary Information for details). Node age ranges were estimated with rates of 0.01 and 0.0276 substitutions per site per lineage per million years (see Supplementary Information). NG, New Guinea; AU, Australia.

pan-Pacific radiation. Given this result and the high degree of morphological and behavioural differentiation within the monarchs¹², estimating the age of this radiation is compelling. The range of divergence dates produced with molecular clock calibrations for island birds¹³ and maximum-likelihood distances from the monarch ND2 data suggest that the Pacific radiation is young (Fig. 1, and Supplementary Information). For example, the initial divergence between island and continental taxa (Fig. 1) is inferred to have occurred within the Pliocene epoch, and the size and plumage differences between Pacific island endemics (Fig. 2), which led to the historic over splitting of the group, might have evolved as recently as the Pleistocene. These apparent rates of morphological evolution are not surprising, given the timing of spectacular radiations within Pacific archipelagos such as the Galapagos and the Hawaiian chain¹⁴. However, our data indicate that rapid geographic dispersion and morphological diversification might also occur at the scale of the entire tropical Pacific.

Regardless of the timing of diversification, the branching pattern in the island clade does not follow the expected stepping-stone pattern from inshore to remote islands. Instead, the most basal division among the island taxa separates eastern and western clades (Fig. 2). Thus, one clade encompasses Micronesia and insular Melanesia, and the other includes all of central and eastern Polynesia. The two meet, but do not overlap, in the eastern Solomon Islands. Exploring a link between this biogeographic pattern and recent geotectonic reconstructions¹⁵ demands comparative systematic analyses of additional groups and finer-grained geological interpretation. In the light of recent archaeological evidence¹⁶,

these biogeographic results must also be considered within the broader context of human-induced extinction that has denuded the Pacific avifauna. In the Pacific, possibly more than any other area on Earth, we are forced to interpret distribution patterns that are incomplete. In monarchs, direct evidence of historic extinctions is sparse, but some pruning of lineages in the Pacific is nearly certain. However, if patterns of extinction follow those in other lineages, where archaeological evidence shows disproportionate anthropogenic extirpation of island taxa in comparison with continental forms¹⁶, accounting for recent extinctions would probably strengthen our interpretation of a species-rich radiation of monarchs in the tropical Pacific.

These results strongly support a recent, rapid sequence of colonization and diversification across all major archipelagos in the Pacific, followed by subsequent recolonization of Australia and New Guinea (Figs 1 and 2). The single-lineage approach employed in our analysis is especially revealing not only because this particular lineage has had historical influence on biogeographic and speciation theory^{2,8,11} but also because islands support only a small subset of continental families. Relatively few families contribute most Pacific bird species, so a few extensive insular radiations would account for a large proportion of diversity. The Pacific monarch radiation itself might be more extensive if it includes the enigmatic Fijian genus *Lamprolia*¹⁷, the only endemic genus not sampled for this study. Prospects of additional lineages in the Pacific having a similar biogeographic history to that of the monarchs seem likely when one considers taxonomic patterns in other bird families. Like the monarchs, pigeons (Columbidae), rails (Rallidae), and white-eyes

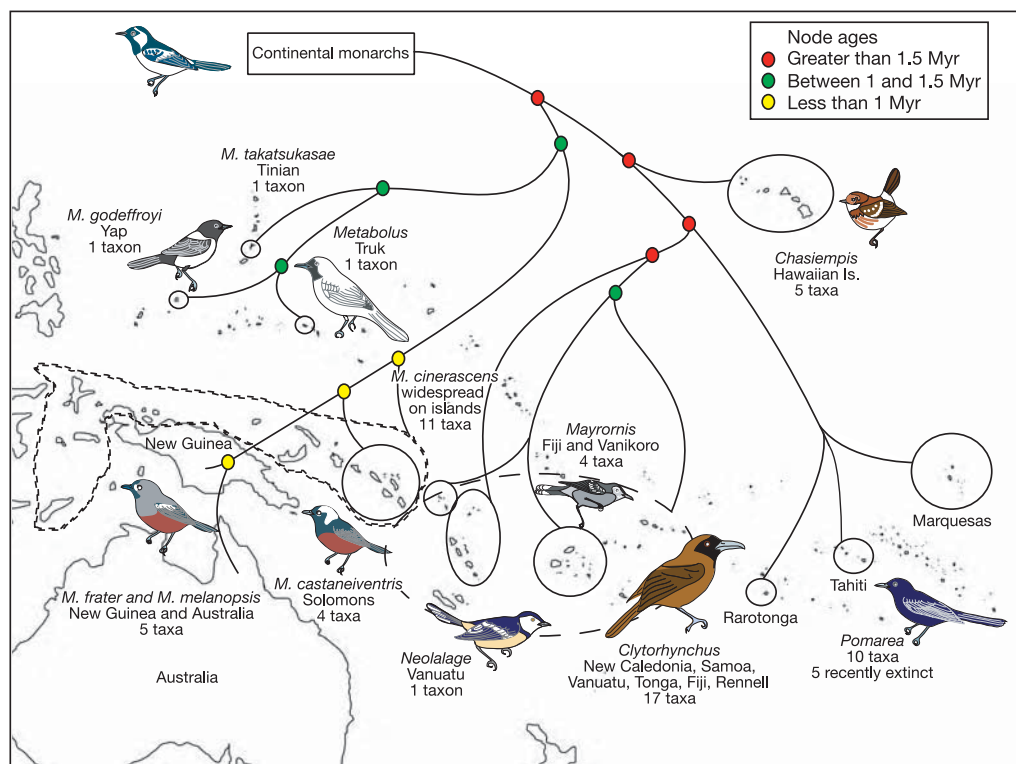


Figure 2 | Evolution of Pacific monarchs (Fig. 1, clade A) in space, time and form. Bird drawings are roughly to scale and illustrate some of the marked morphological differentiation between island endemics that has complicated the evolutionary interpretation of the group. Coloured circles indicate three classes of node age by using point estimates from ND2-specific

rate calibration (see Supplementary Information). Branch lengths are not to scale. The image representing the continental monarchs (*Monarcha verticalis*) is typical of taxa within clade B (Fig. 1). Large hatching outlines the distribution of *Clytorhynchus* and small hatching outlines the range of *M. cinerascens*.

(Zosteropidae) all have Pacific distributions characterized by small endemic genera. Ultimately, Mayr's inability to find a single genus with a Polynesian centre of origin can be attributed to an incomplete knowledge of the relationships between endemic taxa and to his reliance on the standard, but inadequate, taxonomic methods of his day. In the absence of modern phylogenetic approaches, patterns of continental affinity and phylogeny in Pacific endemics were indecipherable.

Previous phylogenetic studies of tropical Pacific birds tended to focus on patterns within single archipelagos^{18–20}. Here we have shown that a systematic focus on the scale of the entire Pacific markedly alters assumptions about the origins of both insular and continental faunas. Similar analyses have suggested Pacific origins for major arthropod radiations²¹ but have not challenged the pervasive assumption of simple one-way continent-to-island colonization. Using an avian group cited to support early biogeographic hypotheses, we have described the first example of small insular avifaunas providing significant input into a large, taxonomically rich system. This study therefore indicates the need for a reassessment of island biogeographic theory. Evolutionary ecologists have increasingly shown that the process of species accumulation on islands is far more complex than previously thought by questioning the reality of an equilibrium number of species²², the homogeneity of immigration filters²³, and species-specific effects on colonization and persistence²⁴. Palaeontological perspectives also indicate the need for a significant revision of theory by providing evidence that historic patterns of species richness are not reflected in extant island avifaunas, particularly in the Pacific²⁵. We have demonstrated the necessity of historical approaches in re-evaluating the general assumptions of regional source-sink dynamics between islands and continents. Upstream colonization from islands to continents may be far more important than previously realized²⁶ and can only be detected by systematic

studies on broad geographic scales. As the monarch radiation in Australasia and the Pacific shows, considering island groups collectively may reveal a more important role of insular faunas in producing global diversity.

METHODS

We sampled species from throughout the monarch flycatchers but focused on those genera in the Pacific region (see Supplementary Information for a list of taxa and vouchers), including samples from all species groups identified by previous molecular work¹². We sequenced the mitochondrial nicotinamide adenine dinucleotide dehydrogenase (ND2) gene and a nuclear intron (myoglobin intron 2) for all samples for which fresh tissue was available. DNA was extracted from museum study skins for some taxa and used to amplify and sequence the ND2 gene only (see Supplementary Information for primer sequences). Sequences were aligned by eye, which was straightforward because of a lack of insertions and deletions (indels) in the mitochondrial data and the extreme rarity of indels in the intron data.

Congruence between the phylogenetic signal in the two genes was tested with the incongruence length difference test²⁷, implemented in Paup*4.0b10 (the partition homogeneity test²⁸). The test excluded constant characters and ran for 1,000 bootstrap repetitions. We analysed the data under maximum-parsimony and maximum-likelihood criteria by using PAUP*4.0b10 (ref. 28). In likelihood analyses, Modeltest 3.6 (ref. 29) determined both the appropriate model of nucleotide substitution and the parameter values. In parsimony analyses all characters were equally weighted. Support for nodes in the resulting phylogenetic hypotheses was assessed by non-parametric bootstrapping, and reanalysis of the resulting data (100 maximum-likelihood replicates, 1,000 maximum-parsimony replicates). Mixed-model bayesian analysis was performed on the combined data with the use of MRBAYES 3.04 (ref. 30) for 10 million generations.

Parsimony and maximum-likelihood methods were used to infer ancestral areas for Pacific monarchs (see Supplementary Information). Species were coded geographically into one of two categories: continental or island. The mitochondrial data were tested for departure from clock-like evolution with a likelihood

ratio test. Divergence dates were estimated by using a rate (0.0276 substitutions per site per lineage per million years) calculated from island ages and *ND2* divergences in Galapagos mockingbirds¹³, as well as a minimum rate (0.01 substitutions per site per lineage per million years) derived from several published avian rate calibrations (see Supplementary Information for a discussion of avian rate calibrations).

Received 26 April; accepted 19 July 2005.

- Wallace, A. R. On the law which has regulated the introduction of new species. *Ann. Mag. Nat. Hist.* **26**(2), 184–196 (1855).
- Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
- Lack, D. *Darwin's Finches* (Cambridge Univ. Press, Cambridge, 1947).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, 1967).
- Diamond, J. M. Colonization of exploded volcanic islands by birds: the supertrap strategy. *Science* **184**, 803–806 (1974).
- Wilson, E. O. The nature of the taxon cycle in a tropical ant fauna. *Evolution* **13**, 122–144 (1961).
- Diamond, J. Continental and insular speciation in Pacific island birds. *Syst. Zool.* **26**, 263–268 (1977).
- Mayr, E. & Diamond, J. M. *The Birds of Northern Melanesia: Speciation, Ecology and Biogeography* (Oxford Univ. Press, New York, 2001).
- Gillespie, R. G. Biogeography of spiders on remote oceanic islands of the Pacific: archipelagoes as stepping stones? *J. Biogeogr.* **29**, 655–662 (2002).
- Wright, S. D., Yong, C. G., Dawson, J. W., Whittaker, D. J. & Gardner, R. C. Riding the ice age El Niño? Pacific biogeography and evolution of *Metrosideros* subg. *Metrosideros* (Myrtaceae) inferred from nuclear ribosomal DNA. *Proc. Natl Acad. Sci. USA* **97**, 4118–4123 (2000).
- Mayr, E. The origin and the history of the bird fauna of Polynesia. *Proc. VI Pac. Sci. Cong.* **4**, 197–216 (1939).
- Filardi, C. E. & Smith, C. E. Molecular phylogenetics of monarch flycatchers (genus *Monarcha*) with emphasis on Solomon endemics. *Mol. Phylogenet. Evol.* published online, 11 March 2005 (doi:10.1016/j.ympev.2005.02.007).
- Drovetski, S. V. et al. Complex biogeographic history of a Holarctic passerine. *Proc. R. Soc. Lond. B* **271**, 545–551 (2004).
- Grant, P. R. Reconstructing the evolution of birds on islands: 100 years of research. *Oikos* **92**, 385–403 (2001).
- Hall, R. in *Biogeography and the Geological Evolution of SE Asia* (eds Hall, R. & Holloway, J. D.) 99–131 (Backhuys, Leiden, 1998).
- Steadman, D. W. Prehistoric extinctions of Pacific island birds; biodiversity meets zooarchaeology. *Science* **267**, 1123–1131 (1995).
- Olson, S. L. *Lamprolia* as part of a South Pacific radiation of monarchine flycatchers. *Notornis* **27**, 7–10 (1980).
- Tarr, C. L. & Fleischer, R. C. Mitochondrial DNA variation and evolutionary relationships in the Amakihi complex. *Auk* **110**, 825–831 (1993).
- Slikas, B., Jones, I. B., Derrickson, S. R. & Fleischer, R. C. Phylogenetic relationships of Micronesian white-eyes based on mitochondrial sequence data. *Auk* **117**, 355–365 (2000).
- Cibois, A., Thibault, J. C. & Pasquet, E. Biogeography of eastern Polynesian monarchs (*Pomarea*): an endemic genus close to extinction. *Condor* **106**, 837–851 (2004).
- Spironello, M. & Brooks, D. R. Dispersal and diversification: macroevolutionary implications of the MacArthur–Wilson model, illustrated by *Simulium* (*Inselliellum*) Rubstov (Diptera: Simuliidae). *J. Biogeogr.* **30**, 1563–1573 (2003).
- Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science* **294**, 1522–1524 (2001).
- Ward, S. A. & Thornton, I. W. B. Chance and determinism in the development of isolated communities. *Global Ecol. Biogeogr.* **9**, 7–18 (2000).
- Lomolino, M. V. Mammalian community structure on islands: immigration, extinction, and interactive effects. *Biol. J. Linn. Soc.* **28**, 1–21 (1986).
- Steadman, D. W. Biogeography of Tongan birds before and after human impact. *Proc. Natl Acad. Sci. USA* **90**, 818–822 (1993).
- Fjeldså, J. The origin of species in Melanesia. *Trends Ecol. Evol.* **17**, 292 (2002).
- Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. Constructing a significance test for incongruence. *Syst. Biol.* **44**, 570–572 (1995).
- Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer Associates, Sunderland, Massachusetts, 2002).
- Posada, D. & Crandall, K. A. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is dedicated to the memory of Ernst Mayr and his lifelong contribution to biology. We thank G. Barrowclough, J. Cracraft, R. Ricklefs, C. E. Smith and D. Steadman for comments. This work was supported in part by the F. M. Chapman Memorial Fund and by the L. C. and L. J. Sanford Funds of the American Museum of Natural History. The research was conducted in the Ambrose Monell Molecular Laboratory and is a contribution of the Lewis B. and Dorothy Cullman Program for Molecular Systematic Studies, a joint initiative of the New York Botanical Gardens and the American Museum of Natural History.

Author Information Nucleotide sequences newly determined here have been deposited in GenBank under the accession numbers DQ084072–DQ084119. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.E.F. (filardi@amnh.org).

LETTERS

A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression

Scott A. Rifkin^{1,2,†}, David Houle³, Junhyong Kim⁴ & Kevin P. White^{1,2}

Mutation is the ultimate source of biological diversity because it generates the variation that fuels evolution¹. Gene expression is the first step by which an organism translates genetic information into developmental change. Here we estimate the rate at which mutation produces new variation in gene expression by measuring transcript abundances across the genome during the onset of metamorphosis in 12 initially identical *Drosophila melanogaster* lines that independently accumulated mutations for 200 generations². We find statistically significant mutational variation for 39% of the genome and a wide range of variability across corresponding genes. As genes are upregulated in development their variability decreases, and as they are downregulated it increases, indicating that developmental context affects the evolution of gene expression. A strong correlation between mutational variance and environmental variance shows that there is the potential for widespread canalization³. By comparing the evolutionary rates that we report here with differences between species^{4,5}, we conclude that gene expression does not evolve according to strictly neutral models. Although spontaneous mutations have the potential to generate abundant variation in gene expression, natural variation is relatively constrained.

The expression level of a gene is a polygenic, dynamic, quantitative trait, and its functional significance can be assessed by studying genetic variation both within^{6,7} and between (for example, ref. 4) populations or species. Such comparisons need to be calibrated with respect to the effects of mutation. This is commonly measured by the mutational variance (V_m): that is, the per-generation increase in the variance of a trait across a population that is due only to mutation.

On the basis of studies in *D. melanogaster*^{2,8}, we estimate that each of our 12 mutation accumulation lines contains around 360 mutations. We measured gene expression levels during the third larval instar (before puparium formation; BPF) and at puparium formation (PF), before and after the peak of a large pulse of the hormone 20-hydroxyecdysone that triggers the start of metamorphosis (see Methods and Supplementary Fig. 1). This stage is one of substantial transcriptional activity and turnover^{9,10}, with broad intra- and interspecific variation in gene expression⁴. Of 11,798 genes measured, we detected significant V_m for 3,816 genes at the BPF stage, for 3,475 genes at the PF stage and for 4,658 genes overall, using a false discovery rate (FDR) of 0.05. The expression of 5,729 genes significantly differed between the two stages, although only 2,509 of these genes showed significant V_m (FDR = 0.05).

To compare the expression variability of different genes, we scaled estimates of V_m by the residual variance (V_r) to give the mutational heritability (h_m^2) (refs 1, 11). Technical variance is not a significant factor (see Supplementary Methods), so this V_r could arise from inherent physiological variability of expression and/or temporal

asynchrony of the sampled individuals. The 95% interval of h_m^2 for gene expression ranged from 2.7×10^{-6} to 1.2×10^{-4} with medians of 2.5×10^{-5} and 2.3×10^{-5} for BPF and PF, respectively (Fig. 1). These estimates lie at the low end of the variability spectrum, overlapping life-history traits such as viability in *D. melanogaster* and grain yield in barley¹². Although the h_m^2 values were relatively low, the transcript abundances of roughly three-fifths of the 7,878 genes with measurable mRNA levels in these two stages varied among our lines, which is ample material for rapid evolutionary change (Supplementary Data 1).

Some authors have claimed that gene expression evolves neutrally on the basis of correlations between sequence and expression divergence¹³ or on the basis of functional divergence of orthologous genes between humans and mice¹⁴. Other studies suggest that stabilizing selection has a more important role^{4,15–17}. The expected neutral divergence between species depends on the baseline mutation rate for gene expression, which we can derive from the estimates of V_m in the mutation accumulation lines (see Methods)¹⁸. Species divergences significantly smaller than this expectation would be consistent with either stabilizing selection within each species towards a shared (ancestral) value or other kinds of constraint that prevent neutral divergence.

Under mutation-drift equilibrium, in the absence of selection, the expected difference between phenotypic values in two lineages is $\sqrt{(2tV_m)}$, where t is the number of generations since their common ancestor¹⁸. Using our estimates for V_m , we calculated the expected differences in changes in expression levels from BPF to PF between

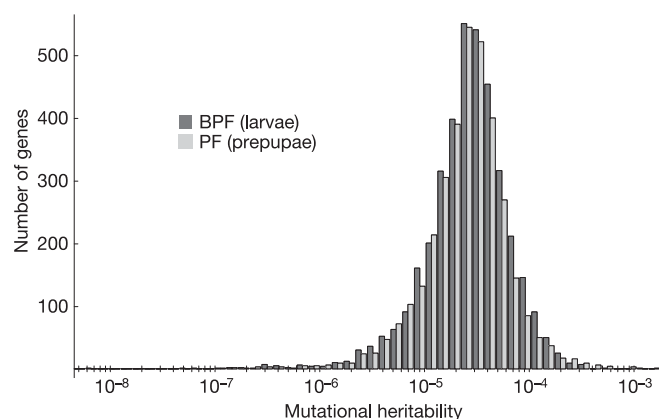


Figure 1 | Mutational heritability. Shown are histograms of the mutational heritability (h_m^2) at the two stages BPF and PF.

¹Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, Connecticut 06520-8106, USA. ²Department of Genetics, Yale University School of Medicine, PO Box 208005, New Haven, Connecticut 06520-8005, USA. ³Department of Biological Science, Florida State University, Tallahassee, Florida 32306-1100, USA.

⁴Department of Biology, University of Pennsylvania, 203 Goddard Laboratories, 415 South University Avenue, Philadelphia, Pennsylvania 19104, USA. [†]Present address: Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

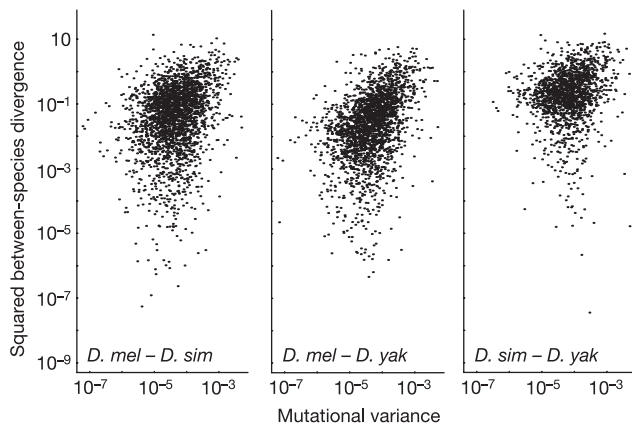


Figure 2 | Mutational variance and differences between species. Shown are log-log plots of V_m of the developmental change between BPF and PF (x axis) versus the squared difference of this change between species (y axis). Measurements of *D. melanogaster* used in the between-species comparisons are from the data in this study, as are the V_m estimates. Measurements of *D. simulans* and *D. yakuba* are from an extended version of published data^{4,5} (Supplementary Methods). r_s for genes with significant V_m in either stage: 0.27 (*D. melanogaster* versus *D. simulans*), 0.45 (*D. melanogaster* versus *D. yakuba*), 0.22 (*D. yakuba* versus *D. simulans*).

D. melanogaster, *Drosophila simulans* and *Drosophila yakuba* on the basis of the divergence times of 2.3 (*D. melanogaster*–*D. simulans*) and 5.1 (*D. melanogaster/simulans*–*D. yakuba*) million years with 10 generations per year^{16,18}. The observed differences between these species^{4,5} (Supplementary Methods) were less than expected (Methods) for almost all genes for which we could measure significant V_m , suggesting that stabilizing selection has a larger role than drift in shaping the evolution of gene expression.

Genes with higher V_m tended, however, to have larger between-species differences (Fig. 2), and patterns of variation across functional groups were also similar between the mutation accumulation and interspecific data. Genes encoding transcriptional regulators had significantly low variability in both stages (multiple test, corrected $P < 10^{-4}$), and the variability of enzymes and structural molecules was high (corrected $P = 0.12$ and 0.058 , respectively, at PF). Between species, genes encoding transcriptional regulators and signal transducers were far less variable in their differential expression than those encoding enzymes and structural proteins^{4,16}. Despite the overall pattern of stabilizing selection, greater mutational input could drive interspecific variation to be higher for some genes than for others. Alternatively, for genes for which changes in expression are deleterious, stabilizing selection may reduce the phenotypic effects of perturbations, canalizing the expression and thereby restricting the potential for gene expression evolution.

Stabilizing selection is most effective at driving canalization when phenotypic variation is environmentally induced³. At both developmental stages, the correlations between V_r and V_m for gene expression are high (Spearman's rank correlation coefficient, r_s : 0.75 at BPF, 0.74 at PF; Fig. 3). If environmental and genetic perturbations affect the mechanisms of gene expression in similar ways¹⁹, for example by affecting the reaction kinetics between transcription factors and binding sites, then organisms probably have the same genetic basis for coping with them²⁰. Although a high correlation between V_m and V_r does not necessarily lead to canalization^{12,21,22}, the evidence for stabilizing selection makes it likely that canalization of gene expression would evolve.

Overall, expression levels and h_m^2 are slightly positively correlated ($r_s = 0.11$; Supplementary Fig. 2). When developmental information is taken into account, however, h_m^2 is inversely related to levels of expression. Among the subset of genes with different h_m^2 in the two stages, those with increasing expression (higher at PF than at

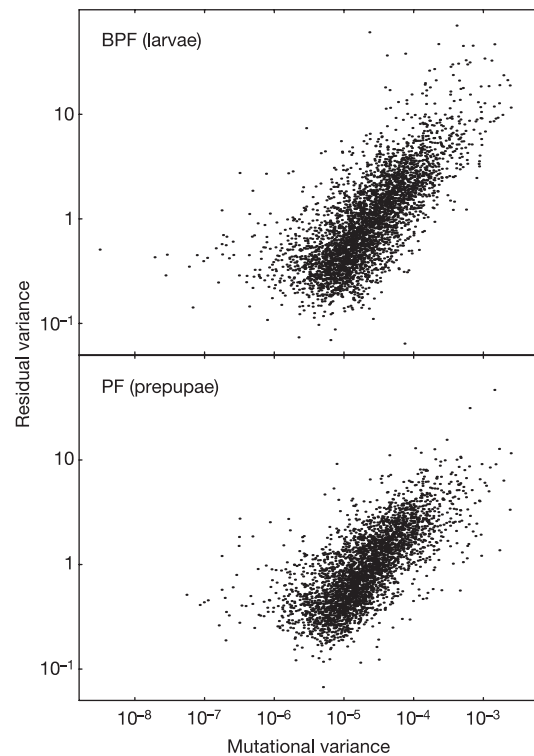


Figure 3 | Opportunity for canalization. Shown are log-log plots of V_m versus V_r . Top, BPF; $r_s = 0.75$. Bottom, PF; $r_s = 0.74$.

BPF) had higher h_m^2 at the earlier stage (median $h_m^2 = 6.2 \times 10^{-5}$) than at the later stage (median $h_m^2 = 2.0 \times 10^{-5}$). Conversely, those with decreasing expression had lower h_m^2 at the earlier stage (median $h_m^2 = 1.4 \times 10^{-5}$) than at the later stage (median $h_m^2 = 4.9 \times 10^{-5}$) ($r_s = -0.55$; Fig. 4). As in interspecific comparisons⁴, as genes are upregulated, their expression becomes more stable and less susceptible to mutations. As genes are downregulated or their transcripts degrade (indicating that large amounts of the proteins that they encode are no longer important for cellular function), variability of expression increases.

At least three-fifths of the genes expressed during these developmental stages vary significantly in these lines (FDR = 0.05), which is as many as the estimated number of mutations. Such widespread change may be due to a few mutations with broadly pleiotropic influences. However, the estimated line effects cluster into about 250 distinct patterns of expression (see Supplementary Methods). The sizes of these clusters are highly skewed, suggesting that some of the variability is due to pleiotropic mutations, but the number of distinct patterns indicates that there is the potential for substantial freedom in the directions of evolution.

Although mutational heritabilities of gene expression are an order of magnitude lower than those of many other traits^{1,22}, between population differences, although substantial⁴, are still far lower than expected under neutral models. There are at least three possible explanations for this restriction. First, expression levels are bounded traits, both above and below. Physical limits on the transcription and degradation rates of mRNA could make it impossible to change expression enough to meet the neutral expectation. Over millions of generations, levels of gene expression evolving at the rates that we report here would rebound off these constraints, erasing any correlation between mutational variance and between-population differentiation. Second, stabilizing selection may act directly on the expression of each gene individually. The sheer number of genes varying along with the complexity of gene interactions in pathways and networks makes this unlikely to be a general explanation. Third,

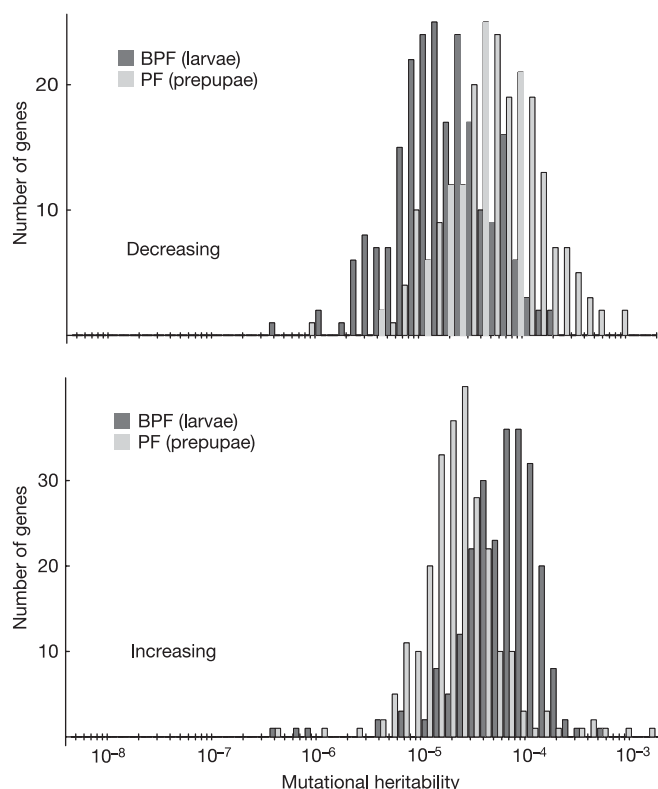


Figure 4 | Developmental context of mutational heritability. Shown are histograms of the h_m^2 of genes with significant but unequal h_m^2 in both stages (FDR = 0.05). Top, genes with significantly decreasing expression between the two stages; bottom, genes with significantly increasing expression between the two stages (FDR = 0.05).

network output, namely the production of a particular product at a specific place and time, may be the target of selection rather than gene expression itself. As in enzyme flux models²³, the selective effect of any particular change in gene expression may be negligible over a range of values but become substantial when the abundance of mRNA becomes rate limiting or when the variation becomes otherwise functionally relevant. Stabilizing selection, by canalizing network output against perturbations, may facilitate neutrality among members of the underlying network²⁴. Gene expression would be able to tolerate a moderate number of mild mutations but would trigger strong selection if the network output were substantially affected. Such a model could also account for moderate correlations between mutational and interspecific variation even when the total level of between-species divergence is far less than expected under neutrality.

In summary, *D. melanogaster* has a broad mutational capacity for changes in gene expression, in both magnitude and genomic extent, that could potentially provide ample raw material for evolutionary diversification. However, although they vary among closely related species, gene expression patterns are relatively stable⁴. In *Caenorhabditis elegans*, genetic variances of gene expression are likewise much less than the neutral expectation¹⁷. The convergence of this observation in two groups of organisms that diverged in the Precambrian²⁵ and have different reproductive and life-history strategies indicates that stabilizing selection and structural processes, including canalization, physical and developmental constraints, and correlated responses, govern gene expression evolution.

METHODS

Flies. The 12 mutation accumulation lines were randomly chosen from surviving *D. melanogaster* Ives strain (Ive-39) sub-lines described in ref. 2, after about 200 generations of mutation accumulation (Supplementary Methods). Immediately before sampling, they were rapidly expanded over not more than four

generations in uncrowded conditions on standard cornmeal medium containing 0.05% bromophenol blue and sprinkled dry yeast. About 18 h before pupariation, we collected flies with dark full guts (BPF) at the stage when they start to crawl up the side of the bottle to pupariate. For the puparium formation stage, we cleared the bottles of pupae and picked newly pupariated flies in 30-min windows (Supplementary Fig. 1).

Microarray hybridizations. We designed the hybridizations to efficiently estimate stage-specific across-line and within-line variance (Supplementary Methods and Supplementary Fig. 3). Each of the 12 lines was measured eight times at each of two stages. We ground about 30 flies in TRIzol (Invitrogen), extracted total RNA by adding chloroform, extracted mRNA with oligo-dT cellulose (Ambion) and poly-prep columns (BioRad), and prepared labelled sample using the Powerscript fluorescent labelling kit (BD Biosciences) and monofunctional Cy3 and Cy5 dyes (Amersham). We printed and processed whole-genome *D. melanogaster* microarray slides as described⁴, and hybridized and washed the samples according to a slightly modified version of protocol M005 from The Institute for Genomic Research (<http://www.tigr.org>). We scanned the slides with a GenePix 4000 series scanner (Axon) and analysed the images using Spot 2.0 (CSIRO) with modifications for manually flagging bad spots. We carried out a series of global and array-specific normalizations to remove dye-, intensity-, location-, and scale-dependent effects (Supplementary Methods).

Quantitative genetic analyses. After the global and array-specific normalizations, we fit the data for each gene individually to several linear mixed models using Python with calls to PROC MIXED^{26–28} in SAS software v.8.2 (SAS Institute; Supplementary Data 2). Roughly 1% of the measurements in the data set were flagged and not used. Because PROC MIXED uses restricted maximum likelihood, missing data were not the problem that they would be in a moments-based estimation.

The gene-specific full model, allowing for stage-specific mean effects, sequence effects when several spots represent the same gene, spot effects, stage-specific across line variances and stage-specific residuals is

$$y_{ijkq} = \mu + \text{Stage}_i + \text{Sequence}_q + \text{Array}(\text{Sequence})_{j(q)} + \text{Line}(\text{Stage})_{k(i)} + \varepsilon_{ijk} \quad (1)$$

$$\text{Array}(\text{Sequence}) \sim N(0, \sigma_{A(q)}^2); \text{Line}(\text{Stage}) \sim N(0, \sigma_{L(i)}^2); \varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon(i)}^2)$$

where y_{ijkq} is the $\log_2$ measurement for a sequence q for a particular gene at stage i from line k on array j , μ denotes the grand mean, and $N(0, \sigma^2)$ means the normal distribution with variance σ^2 , respectively subscripted for each component. Specifically, $\sigma_{A(q)}^2$ is the variance for a particular spot across arrays, and $\sigma_{L(i)}^2$ represents the variance across lines at stage i . The overall mean, stage and sequence effects are fixed; the array, line and error effects are random. Preliminary analyses with models including a covariance term could not estimate significant correlation between stages, possibly because of insufficient power. Because each of our samples contained about 30 flies, we multiplied the estimate of $\sigma_{\varepsilon(i)}^2$ from our model by 30, and used this residual variance (V_m) as a surrogate for the environmental variance. If there is appreciable intra-sample covariance within a line, this correction will overestimate the environmental variance. We estimated V_m as 1/400 of the across-line variance $\sigma_{L(i)}^2$ (ref. 18). On the basis of a power analysis, we would detect an across-line variance of 0.02 with a probability of 0.9 given the median within-line variance of 0.023.

We constructed a hierarchy of six alternative models that relaxed the assumptions of stage-specificity of σ_L^2 , stage-specificity of σ_{ε}^2 , and the existence of line-specific effects (Line_k). Using a series of likelihood ratio tests, we determined which model best fit the data and used parameter estimates from that model for further analyses (Supplementary Fig. 4 and Supplementary Methods). We used a false discovery rate²⁹ of 0.05 to account for multiple testing (Supplementary Methods). To reduce bias in our estimates, we subjected the variance estimates to a jack-knife procedure (Supplementary Methods)³⁰.

Comparisons to interspecific differences used an updated version of published data^{4,5} (Supplementary Methods). We tested for divergence from neutrality using a two-tailed $F_{(1,11)}$ -test of the ratio of squared divergence between species to the neutral expectation (FDR = 0.05). A Kruskal–Wallis test showed that functional classes of genes did not significantly differ in the ratio of V_m to squared between-species differences.

Received 26 April; accepted 2 August 2005.

- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1998).
- Houle, D. & Nuzhdin, S. V. Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. *Genet. Res.* 83, 7–18 (2004).

3. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51**, 329–347 (1997).
4. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
5. Gu, Z. L., Rifkin, S. A., White, K. P. & Li, W. H. Duplicate genes increase gene expression diversity within and between species. *Nature Genet.* **36**, 577–579 (2004).
6. Oleksiak, M. F., Roach, J. L. & Crawford, D. L. Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genet.* **37**, 67–72 (2004).
7. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
8. Yang, H.-P., Tanikawa, A. Y. & Kondrashov, A. S. Molecular nature of 11 spontaneous *de novo* mutations in *Drosophila melanogaster*. *Genetics* **157**, 1285–1292 (2001).
9. White, K. P., Rifkin, S. A., Hurban, P. & Hogness, D. S. Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**, 2179–2184 (1999).
10. Li, T.-R. & White, K. P. Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. *Dev. Cell* **5**, 59–72 (2003).
11. Lande, R. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* **26**, 221–235 (1976).
12. Houle, D., Morikawa, B. & Lynch, M. Comparing mutational variabilities. *Genetics* **143**, 1467–1483 (1996).
13. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, e132 (2004).
14. Yanai, I., Graur, D. & Ophir, R. Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *OMICS* **8**, 15–24 (2004).
15. Hsieh, W.-P., Chub, T.-M., Wolfinger, R. D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).
16. Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137 (2005).
17. Denver, D. R. *et al.* The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genet.* **37**, 544–548 (2005).
18. Lynch, M. & Hill, W. G. Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935 (1986).
19. Waddington, C. H. Genetic assimilation of an acquired character. *Evolution* **7**, 118–126 (1953).
20. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).
21. Houle, D. How should we explain variation in the genetic variance of traits? *Genetica* **103**, 241–253 (1998).
22. Lynch, M. *et al.* Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).
23. Kacser, H. & Burns, J. The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973).
24. Hartl, D., Dykhuizen, D. & Dean, A. Limits of adaptation: the evolution of selective neutrality. *Genetics* **111**, 655–674 (1985).
25. Benton, M. J. & Ayala, F. J. Dating the tree of life. *Science* **300**, 1698–1700 (2003).
26. Wolfinger, R. *et al.* Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.* **8**, 625–637 (2001).
27. Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. *SAS System for Mixed Models* (SAS Institute, Cary, North Carolina, 1996).
28. Fry, J. D. in *Genetic Analysis of Complex Traits Using SAS* (ed. Saxton, A.) 11–34 (SAS Institute, Cary, North Carolina, 2004).
29. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
30. Sokal, R. R. & Rohlf, F. J. *Biometry* (W. H. Freeman and Company, New York, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Wagner, S. Rice, J. Powell, M. Lynch, J. Fry, G. Gibson, P. Gibbs, T.-R. Li, J. D. Lambert and members of the Kim and White laboratories for suggestions, help and comments. This work was supported by a National Library of Medicine fellowship (to S.A.R.), an NIH grant (to J.K.), and grants from the W.M. Keck Foundation, the Arnold and Mabel Beckman Foundation, and the NIH and National Human Genome Research Institute (to K.P.W.).

Author Contributions S.A.R. and J.K. planned and designed the project in consultation with D.H. and K.P.W. Expression data was collected by S.A.R. using spotted microarrays developed by K.P.W. D.H. generated and maintained the mutation accumulation lines. S.A.R. and J.K. developed the computational analyses and carried out the quantitative genetics modelling. S.A.R. wrote the paper with contributions from all authors.

Author Information Microarray data have been deposited in the Gene Expression Omnibus under accession numbers GSE2126 (mutation accumulation lines), GSE2641 (technical error) and GSE2642 (comparative data update and extension). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.K. (junhyong@sas.upenn.edu) or K.P.W. (kevin.white@yale.edu).

LETTERS

Gigaxonin-controlled degradation of MAP1B light chain is critical to neuronal survival

Elizabeth Allen^{1*}, Jianqing Ding^{1*}, Wei Wang¹, Suneet Pramanik¹, Jonathan Chou¹, Vincent Yau¹
& Yanmin Yang¹

Giant axonal neuropathy (GAN) is a devastating sensory and motor neuropathy caused by mutations in the *GAN* gene, which encodes the ubiquitously expressed protein gigaxonin^{1–5}. Cytopathological features of GAN include axonal degeneration, with accumulation and aggregation of cytoskeletal components^{6–7}. Little is currently known about the molecular mechanisms underlying this recessive disorder. Here we show that gigaxonin controls protein degradation, and is essential for neuronal function and survival. We present evidence that gigaxonin binds to the ubiquitin-activating enzyme E1 through its amino-terminal BTB domain, while the carboxy-terminal kelch repeat domain interacts directly with the light chain (LC) of microtubule-associated protein 1B (MAP1B)⁸. Overexpression of gigaxonin leads to enhanced degradation of MAP1B-LC, which can be antagonized by proteasome inhibitors. Ablation of gigaxonin causes a substantial accumulation of MAP1B-LC in GAN-null neurons. Moreover, we show that overexpression of MAP1B in wild-type cortical neurons leads to cell death characteristic of GAN-null neurons, whereas reducing MAP1B levels significantly improves the survival rate of null neurons. Our results identify gigaxonin as a ubiquitin scaffolding protein that controls MAP1B-LC degradation, and provide insight into the molecular mechanisms underlying human neurodegenerative disorders.

Gigaxonin is a distant member of the BTB/kelch superfamily^{1,5}. The devastating degeneration and cell death occurring in GAN patients point to the importance of gigaxonin for axonal structure and neuronal maintenance. We have previously characterized the direct association between the kelch repeat domain of gigaxonin and the light chain of MAP1B⁸. Recent studies indicate that BTB proteins might function as substrate-specific adaptors, forming a new class of SCF-like (Skp1-Cul1-Fbox) ubiquitin ligases^{9–11}. These proteins complex with a cognate E2 and the ubiquitin-activating enzyme E1 (E1) to direct ubiquitin-mediated protein degradation. Although recombinant gigaxonin has been shown to co-immunoprecipitate with Cul3 from transfected cells¹², there is no evidence linking gigaxonin and the ubiquitin–proteasome system to the pathogenesis of GAN.

To elucidate the role of gigaxonin in specific biological pathways, a yeast two-hybrid screen was used to identify its interactors. Multiple positive clones pointed to the ubiquitin-binding domain of E1 (amino acids 530–638) as a binding partner of gigaxonin (Fig. 1a). To define the interaction site on gigaxonin, we divided the protein into BTB and kelch repeat domains for further mapping in yeast, and found that the BTB motif binds E1 (Fig. 1b). Additional evidence for the direct interaction between E1 and gigaxonin-BTB was obtained in pull-down assays using purified proteins. Immobilized His-BTB domain from wild-type gigaxonin was incubated with a purified glutathione S-transferase (GST)–E1 fusion protein or an irrelevant

(control) GST fusion protein. In immunoblot assays, wild-type His-BTB bound only to GST–E1 (Fig. 1c, lanes 1 and 1') but not to control proteins (Fig. 1c, lanes B and B', 3, and 3'). The direct association between the gigaxonin BTB domain and E1 establishes that gigaxonin is linked to the ubiquitin–proteasome system. Notably, the binding between GST–E1 and two different gigaxonin mutants, V82F (Fig. 1c, lanes 2 and 2') and I85F (data not shown), was significantly impaired, showing that GAN mutations can disrupt the gigaxonin–E1 interaction.

The association of gigaxonin with E1 in mammalian cells was verified by co-immunoprecipitation from PC12 cells using a rabbit anti-gigaxonin antibody, followed by immunoblotting with an anti-E1 antibody (Fig. 1d, lane 1). Additional evidence was obtained from pull-down assays using cells transfected with Flag-tagged wild-type or mutant V82F gigaxonin, or an irrelevant Flag-tagged protein. As shown in immunoblot analysis, endogenous E1 protein bound only to immobilized Flag-tagged wild-type gigaxonin, but not to the V82F mutant gigaxonin or control protein (Fig. 1d, lanes 3–5), further demonstrating the V82F point mutation in gigaxonin impairs its specific interaction with E1.

Defects in ubiquitin–proteasome system function have recently been linked to axonal degeneration^{13–15}. Finding an association between gigaxonin and E1 prompted us to investigate whether gigaxonin has a role in controlling the degradation of MAP1B-LC through a ubiquitin-mediated mechanism. Co-transfection of ubiquitin and MAP1B-LC in PC12 cells resulted in the rapid degradation of Flag–MAP1B-LC in the absence of proteasome inhibitors (Fig. 2a, lane 2). In the presence of proteasome inhibitors, Flag–MAP1B-LC formed an apparently typical ubiquitin ladder (Fig. 2a, lane 3) equivalent to the conjugation of polyubiquitin tags (Fig. 2a, lane 3'). The presence of both Flag–MAP1B-LC and haemagglutinin (HA)–ubiquitin in the high-molecular-mass complexes revealed by co-immunoprecipitations on co-transfected cells further confirms that MAP1B-LC is targeted by the ubiquitin–proteasome system (data not shown).

Notably, when the sequences responsible for binding to gigaxonin were deleted, the N-terminal fragment of MAP1B-LC⁸ was not ubiquitinated (Fig. 2a, lane 5), but smears of HA–ubiquitin indicate the successful ubiquitination of other cellular proteins (Fig. 2a, lane 5'). In the presence of excess Flag-tagged ubiquitin, endogenous MAP1B-LC is consistently ubiquitinated, showing characteristic high-molecular-mass complexes (Fig. 2b, lane 2). The negative result obtained from a duplicate blot probed with anti-actin and anti-dynein intermediate chain (DIC) antibodies shows the specificity of the reaction (Fig. 2b, lanes 1' and 2').

To further characterize MAP1B-LC degradation, we used cyclohexamide treatment to inhibit protein synthesis. We found that less than half of the wild-type Flag–MAP1B-LC protein remains 6 h after

¹Department of Neurology and Neurological Sciences, Stanford University School of Medicine, 1201 Welch Road, Stanford, California 94305-5489, USA.

*These authors contributed equally to this work.

cyclohexamide treatment (Fig. 2c, lanes 1–4). Moreover, Flag-MAP1B-LC levels remain consistent when both proteasome inhibitors and cyclohexamide are present (Fig. 2c, lanes 5–8), providing additional evidence for the requirement of the ubiquitin–proteasome system for degradation. Similar to previous published results¹⁶, endogenous MAP1B-LC protein seems fairly stable, as approximately half of the endogenous protein remained 24 h after cyclohexamide addition (Fig. 2d, lanes 1–6).

To examine whether protein degradation of MAP1B-LC is controlled by gigaxonin, we transiently overexpressed Flag-gigaxonin in PC12 cells. Gigaxonin overexpression significantly reduces MAP1B-LC levels (Fig. 2e, lanes 2, 4, and Supplementary Fig. 1), whereas overexpression of green fluorescent protein (GFP) has no such effect (Fig. 2e, lanes 1, 6, and Supplementary Fig. 1). The effect of gigaxonin was antagonized by the addition of proteasome inhibitors to gigaxonin-overexpressing cells (Fig. 2e, lane 3), in contrast to GFP-overexpressing cells (Fig. 2e, lane 5). Our results demonstrate that gigaxonin has an important role in the degradation of MAP1B-LC, through a ubiquitin-dependent mechanism.

We have established that increased gigaxonin expression leads to enhanced MAP1B-LC degradation. To test the effect of gigaxonin ablation on MAP1B-LC levels, we made total brain or spinal cord

lysates from wild-type or gigaxonin-null mice (J.D., E.A., W.W. & Y.Y., manuscript in preparation). We analysed samples by immunoblotting with antibodies against gigaxonin, actin, DIC and MAP1B-LC. The absence of gigaxonin led to substantially increased MAP1B-LC protein levels in brain and spinal cord (Fig. 3a), whereas levels of DIC and actin remained unchanged. Similar increases in MAP1B-heavy chain (HC) levels have been observed in both brain and spinal cord¹⁷, suggesting that MAP1B might be degraded as a protein complex (data not shown).

To determine whether this increase was due to upregulated gene transcription, we compared messenger RNA levels of mouse *Map1b* (also known as *Mtap1b*) light chain in wild-type and gigaxonin-null animals using polymerase chain reaction with reverse transcription (RT–PCR) and quantitative PCR (Fig. 3b, c). In contrast to the increase in protein levels, RT–PCR and quantitative PCR analyses indicate that transcription of *Map1b* may be slightly downregulated or unchanged in gigaxonin-null mice (Fig. 3b–3c). Thus, increased levels of MAP1B-LC are not due to transcriptional upregulation, but

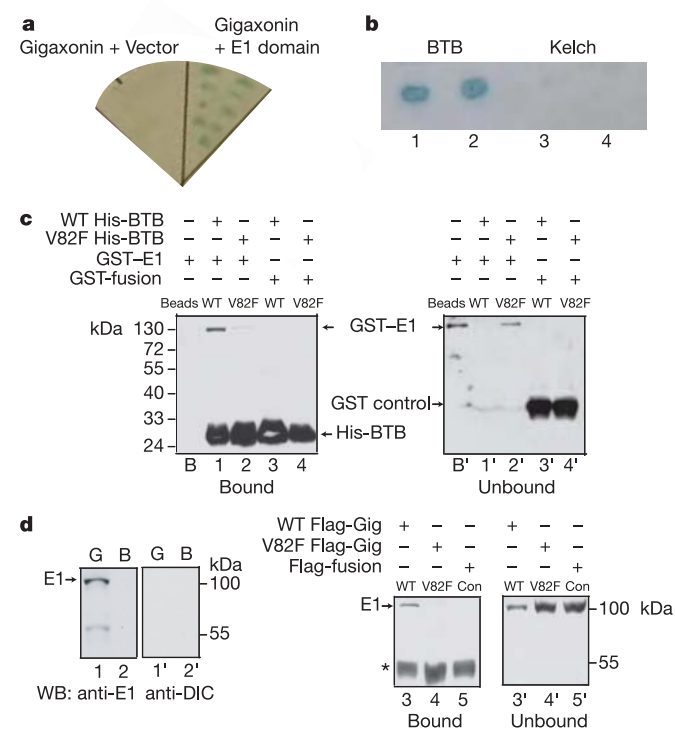


Figure 1 | E1 binds directly to gigaxonin through its BTB domain.

a, pGBT7-gigaxonin (amino acids 11–584) binds pGADT7-E1 (amino acids 530–638) from a human brain library. The negative reaction of pGBT7-gigaxonin with pGADT7 vector control indicates that pGBT7-gigaxonin has no self-transcriptional activity. **b**, Duplicate samples of pGBT7-BTB (amino acids 11–164) (lanes 1, 2) and pGBT7-kelch (amino acids 253–584) (lanes 3, 4) show that the BTB domain of gigaxonin, not the kelch repeat, contains the binding site. **c**, Immobilized wild-type (WT) or V82F His-BTB gigaxonin was incubated with purified GST–E1 or a GST-fusion control, and probed with anti-GST and anti-His antibodies. **d**, Left, co-immunoprecipitation (IP) of PC12 extracts using rabbit anti-gigaxonin antibody (G) or beads only (B, no antibody), followed by immunoblotting with mouse anti-E1 or anti-DIC. Right, Flag-tagged wild-type (WT) or V82F gigaxonin, or an irrelevant Flag-tagged fusion protein, was immobilized on Flag-agarose beads to capture endogenous E1 protein. Bound (bead fraction) and unbound (supernatant) materials were analysed with anti-Flag (not shown) and anti-E1 antibodies. Asterisk, mouse IgG.

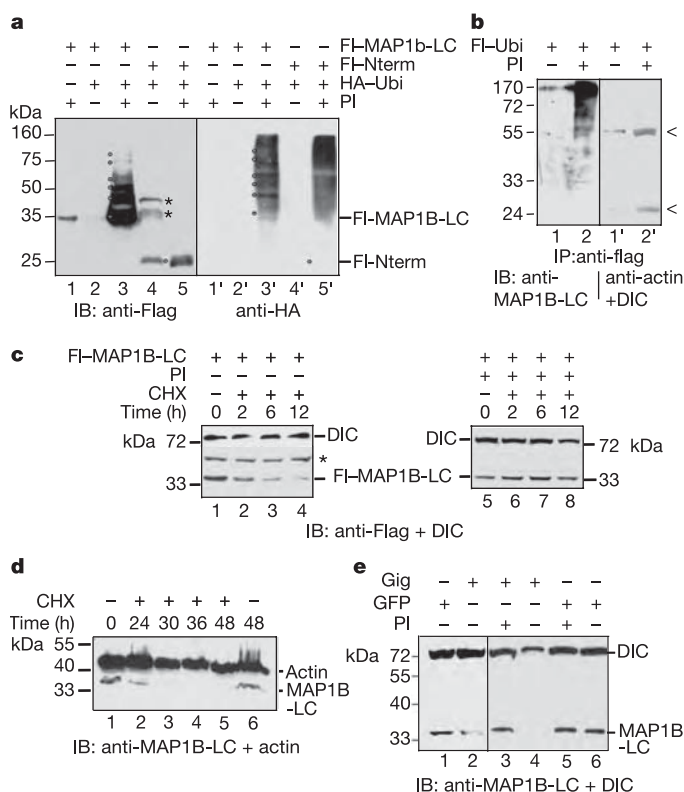


Figure 2 | MAP1B-LC is degraded by a ubiquitin-mediated mechanism.

a, Immunoblots (IB) of PC12 cells co-transfected with HA-ubiquitin (HA-Ubi), Flag-MAP1B-LC or a Flag-tagged N-terminal fragment of MAP1B-LC (FI-Nterm), with or without 10 μ M of the proteasome inhibitor MG132 (PI). Blots were probed with anti-Flag or anti-HA antibody as indicated. Note that in the presence of PI, conjugation of ubiquitin to MAP1B-LC is detected by both anti-Flag and anti-HA (lanes 3 and 3'), whereas no shifting is detected for Flag-tagged N-terminal MAP1B-LC probed with anti-Flag (lanes 5 and 5'). **b**, Ubiquitination of endogenous MAP1B-LC. Flag-tagged ubiquitin-transfected PC12 cells with or without PI were immunoprecipitated with an anti-Flag antibody. Duplicate filters were probed with rabbit anti-MAP1B-LC or mouse anti-actin and anti-DIC. **c**, Stability of exogenous Flag-MAP1B-LC in PC12 cells. Flag-MAP1B-LC-transfected PC12 cells were treated with 50 μ g ml^{−1} cyclohexamide (CHX). **d**, PC12 cells were treated with 10 μ g ml^{−1} CHX to estimate the stability of endogenous MAP1B-LC. **e**, Gigaxonin (Gig) overexpression reduces endogenous MAP1B-LC levels. Cell lysates from gigaxonin- or GFP-transfected PC12 cells in the presence or absence of PI were immunoblotted with anti-MAP1B-LC and anti-DIC antibodies. Asterisk, non-specific band. Arrowhead, IgG.

are caused by a defect in protein degradation, leading to its accumulation. Densitometric scans of western blots indicate that MAP1B-LC levels in 6–8-month-old gigaxonin-null mice are nearly doubled in the brain (Supplementary Fig. 2). Given that the protein was detected only from ~70% of the surviving neurons in gigaxonin-null animals, the actual accumulation could be more dramatic. Consistent with the immunoblot analysis, substantial MAP1B-LC accumulation in gigaxonin-null animals was corroborated by immunohistochemical analysis (Fig. 3e, g, h) compared with wild-type tissues (Fig. 3d, f). More importantly, the number of viable neurons in the hippocampus and dorsal root ganglia of gigaxonin-null mice seemed reduced, and the surviving neurons were appeared disorganized (compare Fig. 4d and e) and morphologically abnormal (Fig. 3e, g, h).

To monitor the neurodegenerative process, we cultured cortical neurons isolated from wild-type or gigaxonin-null mice. After an initial 4–6 days *in vitro*, abundant neurites were seen forming a dense neurite network indistinguishable between wild-type and gigaxonin-null neurons (data not shown). However, neurites of the gigaxonin-null neurons became progressively shorter and sparser after approximately 7–12 days, whereas wild-type neurites continuously grew denser and longer in the following weeks (Fig. 4a). Significant neuronal death occurred in the cultured gigaxonin-null neurons, leading to a marked reduction (up to 90%) in cell density by day 15 (Fig. 4b and Supplementary Fig. 3). This was also supported by abnormal DAPI staining of the remnants of numerous dead cells in the gigaxonin-null cultures (Fig. 4b).

The cell death observed in cultured gigaxonin-null neurons mimics that of neurons *in vivo*. To elucidate whether the neuro-

degeneration and cell death occurring in gigaxonin-null neurons can be attributed to MAP1B-LC accumulation, we overexpressed the MAP1B-LC in wild-type neurons using transient transfection after 3–5 days in *in vitro* culture. Within 24–48 h of transfection, no significant difference was observed between GFP- and MAP1B-LC-transfected cells. However, filamentous structures that are apparently independent of the microtubule network (Fig. 4d) and actin (not shown) were seen in a subset of the MAP1B-LC-overexpressing neurons 48–72 h after transfection (5–20% of transfected neurons in seven independent experiments). By day 5–7 after transfection, approximately 70% of MAP1B-LC-overexpressing neurons began to display fragmented neurites, leading to continuous shortening of neurites until they nearly disappeared. As judged by DAPI staining, all MAP1B-LC-transfected neurons showed extensive nuclear disintegration or condensation 2–4 days after transfection (data not shown). Thereafter, an overwhelming majority completely lost any

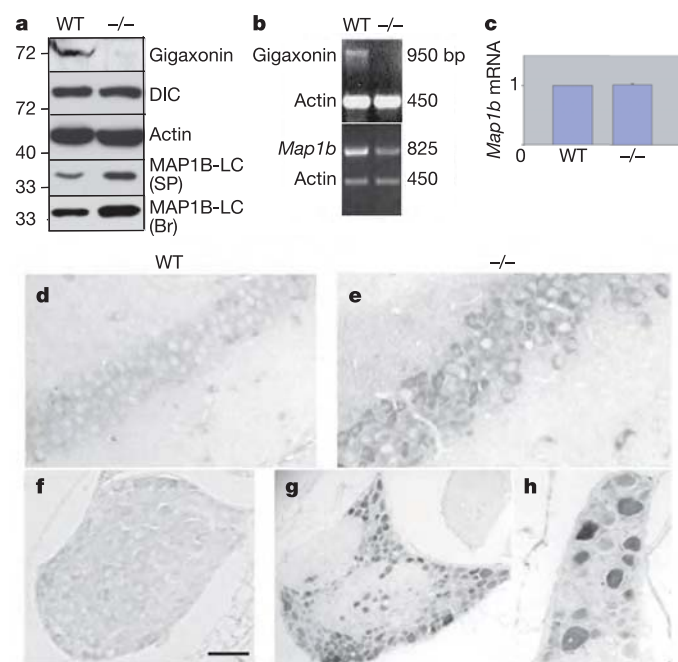


Figure 3 | MAP1B-LC accumulates in gigaxonin-null mice. **a–c**, Protein lysates or mRNAs were prepared from brain (Br) and spinal cord (SP) of 6–9-month-old wild-type (WT) and gigaxonin-null (–/–) mice. **a**, Protein samples were immunoblotted sequentially with antibodies against gigaxonin, DIC, β -actin and MAP1B-LC. **b**, Upper panel, RT-PCR analysis confirms absence of gigaxonin in the gigaxonin-null sample. Lower panel, *Map1b* light-chain transcript level is slightly reduced in gigaxonin-null samples. **c**, Real-time PCR analyses of wild-type and null samples are indistinguishable. Data show means $\pm$ s.d., $P > 0.05$ (Student's *t*-test). **d–h**, Hippocampus (**d**, **e**) and dorsal root ganglia (**f–h**) from wild-type (**d**, **f**) and gigaxonin-null (**e**, **g**, **h**) mice were processed for immunohistochemistry using an anti-MAP1B-LC antibody. Scale bar in **f** represents 20 μ m in **d–f**, 40 μ m in **g**, and 10 μ m in **h**.

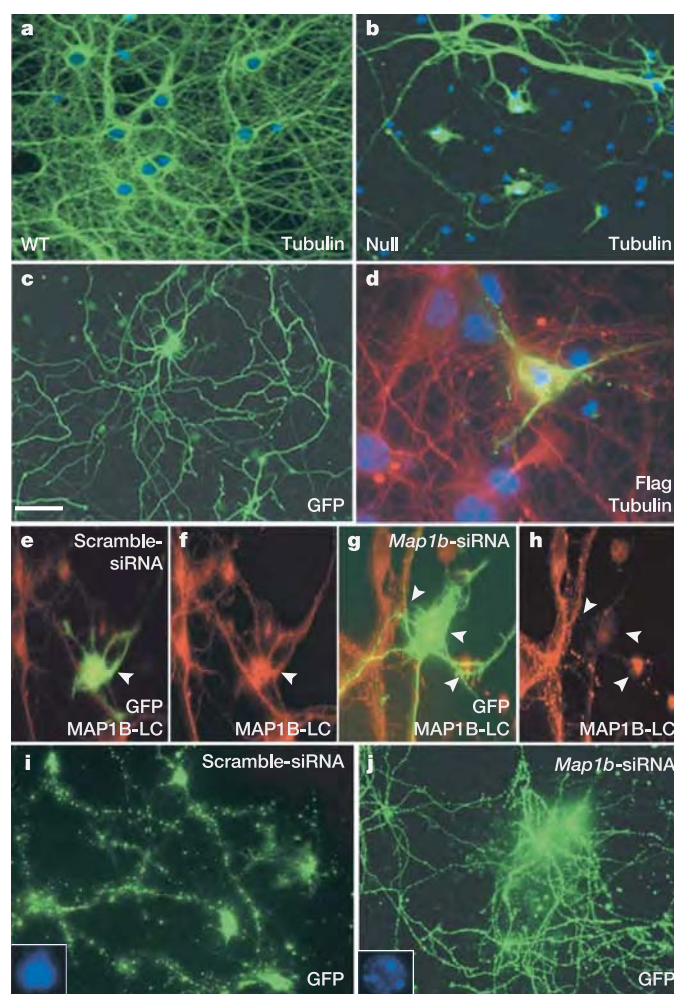


Figure 4 | Cortical neuron cultures from wild-type or gigaxonin-null mice. **a**, **b**, Cortical neurons from wild-type (**a**) and gigaxonin-null (**b**) E15 embryos were cultured for 15 days and co-stained with anti- α -tubulin antibody and DAPI. **c**, **d**, Immunofluorescence of wild-type neurons transfected with GFP (**c**) or Flag-MAP1B-LC (**d**), stained with mouse anti-GFP antibody (**c**) or rabbit anti-Flag antibody (green), mouse anti-tubulin antibody (red) and DAPI (blue) (**d**). **e–h**, Cortical neurons (after 4 days in *in vitro* culture) transfected with scrambled siRNA (**e**, **f**) or *Map1b* siRNA (**g**, **h**) were stained with rabbit anti-MAP1B-LC (red) and mouse anti-GFP (green) antibodies at day 7 in culture. **i**, **j**, Gigaxonin-null cortical neurons transfected with scrambled siRNA (**i**) or *Map1b* siRNA (**j**) after 6 days in culture, stained at day 15 with mouse anti-GFP (green) and rabbit anti-tubulin antibodies (anti-GFP is shown). Scale bar in **c** represents 50 μ m in **a–c**, 75 μ m in **i**, **j**, and 10 μ m in **d–h**.

sign of normal morphology, a phenotype mimicking the degeneration and cell death occurring in cultured gigaxonin-null neurons. In contrast, GFP-expressing and untransfected wild-type neurons were indistinguishable over the entire test period, displaying long axonal processes with marked branching (Fig. 4c).

We have shown that MAP1B-LC overexpression in wild-type neurons is sufficient to cause neuronal degeneration and death. To further assess whether the observed accumulation of MAP1B-LC in gigaxonin-null neurons contributes to neurodegeneration, *Map1b*-specific short interfering RNAs (siRNA) that co-express GFP were transfected into the gigaxonin-null neurons. Immunofluorescence microscopy showed that three days after transfection, *Map1b* siRNA apparently reduces endogenous protein expression (Fig. 4g, h), whereas the scrambled control has no effect (Fig. 4e, f). By day 9–15 after transfection, the *Map1b* siRNA-expressing neurons have a substantially improved survival rate compared with untransfected gigaxonin-null neurons, with numerous elaborate neurite structures and normal DAPI staining (Fig. 4j and Supplementary Fig. 4). In contrast, the scrambled siRNA-expressing neurons are not rescued from neurodegeneration and cell death in gigaxonin-null neurons (Fig. 4i and Supplementary Fig. 4). Thus, we conclude that the accumulation of MAP1B-LC contributes directly to the pathogenesis of neuronal death in GAN. Collectively, our data demonstrate an essential role for gigaxonin in controlling protein levels of MAP1B, and provide direct evidence to link defects in ubiquitin-proteasome system protein degradation with neurodegeneration and cell death.

Several lines of evidence indicate that MAP expression levels are under strict control in response to the morphological and functional requirements of neurons. MAP1B is the first MAP protein expressed during development of the nervous system, and its expression is dramatically downregulated postnatally¹⁸. Recent reports also link mental retardation in fragile X syndrome to a delayed decline in MAP1B expression¹⁹. The fact that MAP1B accumulation leads to progressive cell death further highlights the importance of precisely regulating its expression. Our study defines gigaxonin as a ubiquitin-scaffolding protein that recruits E1 and directs ubiquitin-mediated degradation of MAP1B-LC, a role critical to neuronal maintenance. However, the mechanisms underlying cell death mediated by MAP1B accumulation await future investigation. One possibility is that MAP1B-LC accumulation might interfere with the movement of motor proteins, by directly competing for or blocking adjacent binding sites on the microtubule surface; several studies corroborate this possibility^{20–23}. Another possibility is that aberrant increases in MAP1B levels affect transport processes indirectly by altering microtubule dynamics. Identifying the pathways disrupted by MAP1B-LC accumulation in response defective gigaxonin function may help to illuminate the mechanisms leading to cell death in other human neurodegenerative diseases.

METHODS

GST-E1 and gigaxonin His-BTB binding assays. The BTB domain of wild-type or V82F gigaxonin (amino acids 11–170) was cloned into pET28a to express a His-tagged BTB fusion protein that migrates at approximately 26 kDa. Bacterial lysates were prepared from 50-ml cultures induced with isopropyl- β -D-thiogalactoside (IPTG) to produce 10-ml preparations of soluble material in 1% Triton X-100, 150 mM NaCl, 50 mM Tris (pH 7.4) and protease inhibitor cocktail (Roche Diagnostics). Protein levels were normalized by western blotting. Approximately 6 μ g of wild-type or V82F His-BTB protein was bound to 30 μ l of Ni-NTA beads (Qiagen) for 6 h at 4 °C in 10 mM imidazole, 150 mM NaCl and 100 μ g ml⁻¹ BSA (incubation buffer), and washed with PBS containing 300 mM NaCl and 20 mM imidazole (wash buffer). Purified, active GST-E1 (2.5 μ g, Boston Biochem) or approximately 10 μ g soluble GST-control protein was added to 300 μ l incubation buffer overnight at 4 °C, washed extensively and analysed by immunoblotting with anti-GST and anti-His antibodies. Immunoreactive proteins were visualized using Pico West (Pierce Biotechnology). Immunoprecipitations were performed as previously described⁸.

siRNA experiments. The pZ-OFF EGFP vector was a gift from the laboratory of

C. Garner. The following oligonucleotides for siRNA constructs (Genscript USA) were cloned into the pZ-OFF EGFP vector using *Bgl*II and *Hind*III sites:

MAP1B_forward, 5'-GATCTCGAGATGACCTCACTCTCTTGAATTGATATCCGTTCAAGAGAGTGAAGTGCATCTTTTTTCCAAA-3'; MAP1B_reverse: 5'-AGCTTTTGGAAAAAAGATGACCTCACTCTCTTGAACGGATATCAATTCAAGAGAGTGAAGTGCATCTCGA-3'; Scrambled_forward: 5'-GATCTCGTAGTCTCATGATGGTCAAGCCGTTGATATCCGCGGCTTGACCATCATGAGCTATTTTTCAAA-3'; Scrambled_reverse: 5'-AGCTTTTGGAAAAAATAGTCTCATGATGGTCAAGCCGCGGATATCAACGGCTTGACCATCATGAGCTACGA-3'.

Primary cultures and cortical neuron transfections. Cortical neuron cultures were prepared from mouse embryos at embryonic day (E)12–15, using PCR of tail samples to confirm genotype. Procedures were similar to those described in ref. 24. Briefly, mouse embryo brains were removed and enzymatically dissociated with 0.25% trypsin (Sigma) in dissociation buffer (15 mM HEPES, 1% sodium bicarbonate in 1 × Hanks buffer). Dissociated cells were then plated on 15-mm-diameter coverslips coated with poly-L-lysine (50 μ g ml⁻¹), at a density of 100–160 cells mm⁻². Cortical neurons were then grown in Neurobasal medium containing B27, 0.5 mM L-glutamine, 1% penicillin and 1% streptomycin (100 μ g ml⁻¹) supplements (Gibco). Cultures could be maintained for up to four weeks.

For transfections, cortical neurons were seeded at a density of 1–1.5 × 10⁴ cells per well in a 12-well plate. Expression vectors (Flag-MAP1B-LC or pEGFPN2) were transfected at day 2 or 3 after culturing. siRNA constructs were transfected 4–6 days after culturing and harvested 9–15 days later. Transfection reagents OptiMEM I and Lipofectamine 2000 were used according to the manufacturer's instructions (Invitrogen). All cultures of transfected or untransfected cells were closely observed, counted and compared for cell densities and neurite networks. The cultures were harvested after 3–21 days for immunofluorescence analyses.

Quantitative PCR. Forward and reverse primers were mixed with Syber Green master mix and reference dye (Stratagene) according to the manufacturer's instructions. Data were analysed using the comparative quantification method by the $2^{-\Delta\Delta C_T}$ equation, $\Delta\Delta C_T = \Delta C_T(\text{gigaxonin-null}) - \Delta C_T(\text{wild-type})$, where ΔC_T is the threshold cycle (C_T) value of the housekeeping gene (β -actin) subtracted from the C_T value of the target gene (*Map1b*).

Antibodies and reagents. The rabbit polyclonal anti-MAP1B-LC antibody used on mouse tissue and *in vivo* ubiquitin-shift experiments was a gift from F. Propst²⁵. Mouse anti-MAP1B-HC monoclonal antibodies 3G5 and 6D4 were a gift from L. Binder¹⁷. Mouse anti-MAP1B-LC (M6783, Sigma) was used for PC12 cell studies. Other antibodies included mouse anti-E1 (MAB3520, Chemicon), mouse anti- β -actin (A5441, Sigma), mouse anti-DIC (SC-13524, Santa Cruz), mouse anti- α -tubulin (T9026, Sigma), mouse anti-6 × His (8916-1, BD Biosciences), rabbit anti-GFP (8367-1, BD Biosciences), mouse and rabbit anti-Flag (F3165 and F7425, Sigma), mouse anti-HA (MMMS-101R, Covance) and mouse anti-GST (clone IH12-2, Chemicon). Mouse anti-Flag agarose beads were purchased from Sigma, and purified GST-E1 fusion protein (E-306) was purchased from Boston Biochem.

Received 25 June; accepted 28 September 2005.

Published online 16 October 2005.

1. Bomont, P. *et al.* The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nature Genet.* **26**, 370–374 (2000).
2. Bomont, P. *et al.* Identification of seven novel mutations in the GAN gene. *Hum. Mutat.* **21**, 446 (2003).
3. Tazir, M. *et al.* Genetic heterogeneity in giant axonal neuropathy: an Algerian family not linked to chromosome 16q24.1. *Neuromuscul. Disord.* **12**, 849–852 (2002).
4. Bruno, C. *et al.* Clinical and molecular findings in patients with giant axonal neuropathy (GAN). *Neurology* **62**, 13–16 (2004).
5. Timmerman, V., De Jonghe, P. & Van Broeckhoven, C. Of giant axons and curly hair. *Nature Genet.* **26**, 254–255 (2000).
6. Tandan, R. *et al.* Childhood giant axonal neuropathy. Case report and review of the literature. *J. Neurol. Sci.* **82**, 205–228 (1987).
7. Lampl, Y. *et al.* Giant axonal neuropathy with predominant central nervous system manifestations. *Dev. Med. Child Neurol.* **34**, 164–169 (1992).
8. Ding, J. *et al.* Microtubule-associated protein 1B: a neuronal binding partner for gigaxonin. *J. Cell Biol.* **158**, 427–433 (2002).
9. Pintard, L. *et al.* The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature* **425**, 311–316 (2003).
10. van den Heuvel, S. Protein degradation: CUL-3 and BTB-partners in proteolysis. *Curr. Biol.* **14**, R59–R61 (2004).
11. Geyer, R., Wee, S., Anderson, S., Yates, J. & Wolf, D. A. BTB/POZ domain

- proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. *Mol. Cell* **12**, 783–790 (2003).
12. Furukawa, M., He, Y. J., Borchers, C. & Xiong, Y. Targeting of protein ubiquitination by BTB–Cullin 3–Roc1 ubiquitin ligases. *Nature Cell Biol.* **5**, 1001–1007 (2003).
 13. Ehlers, M. D. Deconstructing the axon: Wallerian degeneration and the ubiquitin–proteasome system. *Trends Neurosci.* **27**, 3–6 (2004).
 14. Zhai, Q. *et al.* Involvement of the ubiquitin–proteasome system in the early stages of wallerian degeneration. *Neuron* **39**, 217–225 (2003).
 15. Watts, R. J., Hoopfer, E. D. & Luo, L. Axon pruning during *Drosophila* metamorphosis: evidence for local degeneration and requirement of the ubiquitin–proteasome system. *Neuron* **38**, 871–885 (2003).
 16. Mei, X., Sweatt, A. J. & Hammarback, J. A. Regulation of microtubule-associated protein 1B (MAP1B) subunit composition. *J. Neurosci. Res.* **62**, 56–64 (2000).
 17. Edelmann, W. *et al.* Neuronal abnormalities in microtubule-associated protein 1B mutant mice. *Proc. Natl Acad. Sci. USA* **93**, 1270–1275 (1996).
 18. Tucker, R. P. & Matus, A. I. Developmental regulation of two microtubule-associated proteins (MAP2 and MAP5) in the embryonic avian retina. *Development* **101**, 535–546 (1987).
 19. Lu, R. *et al.* The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proc. Natl Acad. Sci. USA* **101**, 15201–15206 (2004).
 20. Lopez, L. A. & Sheetz, M. P. Steric inhibition of cytoplasmic dynein and kinesin motility by MAP2. *Cell Motil. Cytoskeleton* **24**, 1–16 (1993).
 21. Paschal, B. M., Obar, R. A. & Vallee, R. B. Interaction of brain cytoplasmic dynein and MAP2 with a common sequence at the C terminus of tubulin. *Nature* **342**, 569–572 (1989).
 22. Hagiwara, H., Yorifuji, H., Sato-Yoshitake, R. & Hirokawa, N. Competition between motor molecules (kinesin and cytoplasmic dynein) and fibrous microtubule-associated proteins in binding to microtubules. *J. Biol. Chem.* **269**, 3581–3589 (1994).
 23. von Massow, A., Mandelkow, E. M. & Mandelkow, E. Interaction between kinesin, microtubules, and microtubule-associated protein 2. *Cell Motil. Cytoskeleton* **14**, 562–571 (1989).
 24. Vicario-Abejón, C. in *Current Protocols in Neuroscience* (eds Crawley, J. N. *et al.*) Ch. 3, 2.1–2.13 (John Wiley & Sons, Somerset, New Jersey, 2004).
 25. Togel, M., Wiche, G. & Propst, F. Novel features of the light chain of microtubule-associated protein MAP1B: microtubule stabilization, self interaction, actin filament binding, and regulation by the heavy chain. *J. Cell Biol.* **143**, 695–707 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Propst and L. Binder for their gifts of rabbit polyclonal MAP1B-LC (ref. 25) and mouse monoclonal MAP1B-HC antibodies, respectively. The pZ-OFF EGFP vector was a gift from C. Garner and P. Zamorano. We also thank C. Wu for helpful discussions. This work was supported by a Muscular Dystrophy Association Research Grant and NIH grants to Y.Y.

Author Contributions E.A. and J.D. contributed equally to this work. E.A. conceived and carried out most of the experiments, with assistance from S.P., J.C. and Y.Y. J.D. conceived and carried out yeast two-hybrid screening, performed quantitative PCR and some biochemical assays, and conducted statistical and pathological analyses. W.W. and J.D. established and maintained cortical neuron cultures. All authors discussed the results. Y.Y. designed and directed the project. E.A. and Y.Y. co-wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.Y. (yyanmin@stanford.edu).

Fruitless specifies sexually dimorphic neural circuitry in the *Drosophila* brain

Ken-Ichi Kimura¹, Manabu Ote², Tatsunori Tazawa¹ & Daisuke Yamamoto^{2,3}

The *Drosophila* *fruitless* (*fru*) gene product Fru has been postulated to be a neural sex determination factor that directs development of the central nervous system (CNS), thereby producing male-typical courtship behaviour and inducing male-specific muscle^{1–6}. Male-specific Fru protein is expressed in small groups of neurons scattered throughout the CNS of male, but not female, *Drosophila*^{4,7}. Collectively, these observations suggest that Fru ‘masculinizes’ certain neurons, thereby establishing neural substrates for male-typical behaviour. However, specific differences between neurons resulting from the presence or absence of Fru are unknown. Previous studies have suggested that Fru might result in sexual differences in the CNS at the functional level, as no overt sexual dimorphism in CNS structure was discernible^{8–10}. Here we identify a subset of *fru*-expressing interneurons in the brain that show marked sexual dimorphism in their number and projection pattern. We also demonstrate that Fru supports the development of neurons with male-specific dendritic fields, which are programmed to die during female development as a result of the absence of Fru. Thus, Fru expression can produce a male-specific neural circuit, probably used during heterosexual courtship, by preventing cell death in identifiable neurons.

Assuming that Fru is involved in development of neural circuits crucial for the generation of male-typical behaviour, it is plausible that Fru-expressing neurons show some degree of sexual dimorphism structurally and/or functionally. Unambiguous identification of Fru-expressing neurons is necessary for evaluating this possibility. In a search for enhancer trap lines that express Gal4 in Fru-expressing cells (see below), we found a conspicuous sex difference in the localization of the Gal4 reporter mouse (m)CD8–GFP (green fluorescent protein) in the *NP21* strain. The *NP21* strain has a P-element insertion in the second intron of the *fru* gene (GETDB database, <http://flymap.lab.nig.ac.jp/%7Edclust/getdb.html>).

The following histological experiments were carried out primarily with flies heterozygous for the described P-element insertion. We observed sexually dimorphic reporter expression in *NP21* flies in two discrete locations: the optic lobe and the region dorsal to the antennal lobe. In the optic lobe, a subset of neurons in the distal medulla expressed mCD8–GFP only in males (Fig. 1a, c). In the region dorsal to the antennal lobe, approximately 30 neurons (30.7 ± 1.7 ; mean $\pm$ s.d., $n = 13$ hemibrains) expressed mCD8–GFP in males, compared with approximately 5 neurons (5.1 ± 0.6 ; $n = 18$) in females (Fig. 1b, d). The somata of these neurons form a cluster called ‘neurons medially located, just above antennal lobe’, or mAL⁷. In the subsequent study, we focus our attention on the mAL cluster.

To identify projection patterns of mAL neurons without interference from surrounding neurons, we used the MARCM method¹¹, which allows mCD8–GFP to be expressed in a single precursor cell and/or its progeny. Somatic chromosomal recombination is induced during

development by the stochastic action of a heat-shock-inducible flippase—by changing the timing of the heat-shock, chromosomal recombination can be induced at different stages of development, resulting in a different number of mCD8–GFP-expressing cells.

Early heat-shock resulted in mCD8–GFP expression in about 30 cells in male flies and 5 cells in female flies, indicating the clonal nature of the mAL cluster. In males, the mAL neurons extended neurite projections bilaterally (Fig. 1e). In females, the neurons only had contralateral projections (Fig. 1f). In both sexes, the contralateral neurite bifurcated after crossing the midline, with one neurite extending dorsally and terminating in the superior lateral protocerebrum, and the other neurite extending ventrally, terminating in the suboesophageal ganglion. The male-specific ipsilateral neurite also terminated in the suboesophageal ganglion. When we carried out single-cell labelling, the 5 cells in females were devoid of this ipsilateral projection (Fig. 1g). We observed a further difference between the sexes in the branching pattern of these neurons in the suboesophageal ganglion: the mAL neurons had forked arborization only in females. The forked arborization typical of female mAL neurons was never observed in males. However, the male mAL had two classes of neurons: one with bilateral projections (Fig. 1h) and the other with only contralateral projections (Fig. 1i). It was possible to differentiate 8 types among male neurons on the basis of their shape (supplementary Fig. S1).

The terminal structures in the superior lateral protocerebrum were decorated with many varicosities (Fig. 1j, k) that were nearly absent from the terminal structures of branches in the suboesophageal ganglion (Fig. 1j, n). This arrangement suggests that the superior lateral protocerebrum is the output site of these interneurons, and the suboesophageal ganglion is the input site. This hypothesis is supported by our observation that a presynaptic marker (synaptotagmin–HA) was transported to the terminals in the superior lateral protocerebrum (Fig. 1k–m) but not to those in the suboesophageal ganglion (Fig. 1n–p).

We then investigated how the timing of somatic recombination by heat-shock application affected the number of labelled neurons (Fig. 1q). When a heat-shock was applied to the embryo, the maximal number of mAL neurons was labelled (that is, about 30 in males and 5 in females; Fig. 1r), indicating that the neuroblast that produced the sexually dimorphic neurons was proliferating in the embryonic stage in both sexes. In contrast, heat-shock activation at the late third instar larval to early pupal stage (that is, 4–5 days after egg collection) failed to label any mAL cells in females, but did result in labelling in males (Fig. 1q). The number of labelled cells in these cases was very low (typically one cell; Fig. 1r), but indicates that the neuroblast continued to proliferate in males at this stage (Fig. 1s, top). These observations are compatible with any of three alternative explanations for the production of fewer neurons in females compared with males (Fig. 1s). First, the female neuroblast stops dividing and

¹Laboratory of Biology, Iwamizawa Campus, Hokkaido University of Education, Iwamizawa 068-8642, Japan. ²Waseda University, Advanced Institute for Science and Engineering, Tokyo 169-8555, Japan. ³Graduate School of Life Sciences, Tohoku University, Sendai 980-8578, Japan.

becomes quiescent during the late third instar larval to early pupal stage. Second, the female neuroblast dies earlier than the male neuroblast. Third, the female neuroblast continues to proliferate as the male counterpart does, but the daughter cells are eliminated by programmed cell death only in females. To differentiate the second

and third possibilities from the first, we examined the effect of mutations that block cell death.

head involution defective (*hid*, also known as *Wrinkled* or *W*), *grim* and *reaper* (*rpr*) are the predominant cell-death genes in *Drosophila*. They are aligned in tandem in the genome, and therefore can be

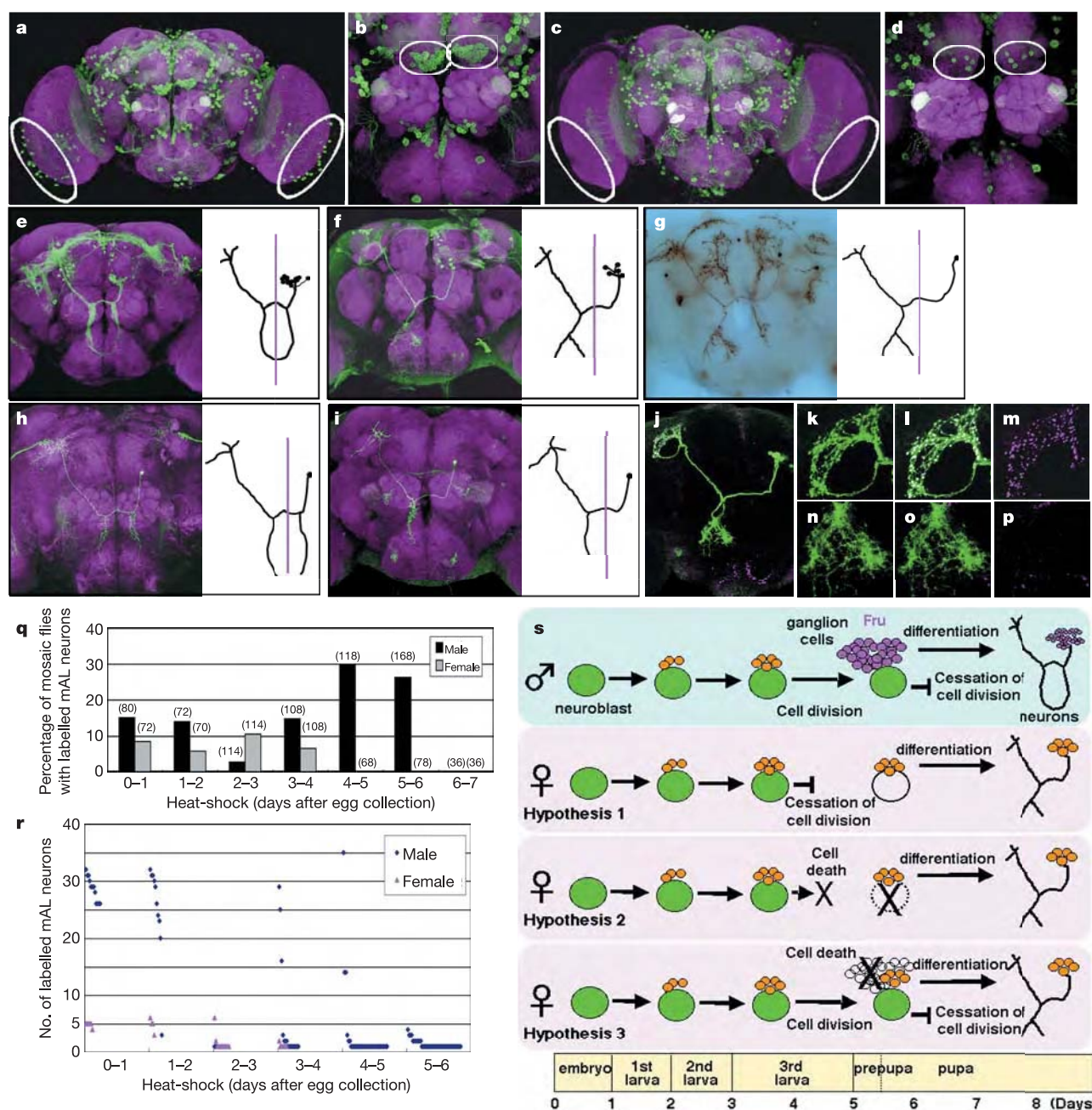


Figure 1 | Sexual dimorphism of mAL neurons in *Drosophila* NP21 strain. **a–d**, Brains of control adult flies (*FRT G13 UAS-mCD8–GFP*; *NP21/TM6B*) doubly stained with an anti-GFP antibody (green) and an anti-nc82 monoclonal antibody (magenta). Regions showing sexual dimorphism are encircled in male (**a, b**) and female (**c, d**) brains. **e–i**, Brain of a mosaic fly with genotype *y hs-flp; FRT G13 UAS-mCD8–GFP/FRT G13 tub-Gal80; NP21/+*. Schematic representations of the projection pattern of mAL neurons are depicted in the right-hand side of each panel. Male (**e**) and female (**f**) brains stained with anti-GFP (green) and anti-nc82 (magenta) antibodies, for flies in which chromosomal recombination was induced during the embryonic stage. Examples of single-cell labelling of mAL neurons in female (**g**) and male (**h, i**) brains. The female brain (**g**) was stained with an anti-GFP antibody and visualized with HRP/DAB. **j**, Merged images of the mAL projection neurites of a female mosaic brain

(*y hs-flp/UAS-synaptotagmin-HA; FRT G13 UAS-mCD8–GFP/FRT G13 tub-Gal80; NP21/+*) stained with anti-GFP (green) and anti-HA (magenta) antibodies. **k–p**, Enlarged images of the mAL cluster arborization in the superior lateral protocerebrum (**k–m**) and the suboesophageal ganglion (**n–p**). Synaptotagmin-HA is localized at varicosities in the superior lateral protocerebrum (**l**, merge) but not in the suboesophageal ganglion (**o**, merge). **q**, The effect of the timing of heat-shock application on the frequency of obtaining mosaic flies with labelled mAL neurons. Numbers in parentheses indicate the number of brain hemispheres examined. **r**, The effect of the timing of heat-shock application on the number of labelled mAL neurons. Each point represents the number of labelled mAL cells counted in a single hemibrain. **s**, Possible mechanisms underlying the formation of sexual dimorphism in the mAL cluster. See the text for details.

removed by a single deficiency, *Df(3L)H99* (ref. 12). In viable *rpr* mutants of *Df(3L)H99/Df(3L)XR38* (ref. 13), the mAL cluster had, on average, 10 mAL neurons in females and 30 in males (Fig. 2a, b). This result is consistent with the idea that, in females, *rpr*-mediated cell death eliminates some cells that otherwise contribute to the mAL cluster. In contrast, viable *hid^{A206}/Df(3L)H99* mutants¹⁴ had approximately 5 mAL neurons in females and 30 in males (Fig. 2a). To observe the effect of removing *hid*, *grim* and *rpr* simultaneously, cells homozygous for *Df(3L)H99* were generated in a heterozygous background by the MARCM method (a homozygous *Df(3L)H99* deficiency causes lethality). Now the mutant females had as many as 29 mAL neurons (mean = 19; Fig. 2a, c). Again, the number of mCD8–GFP-positive cells in mutant males was about 30 (Fig. 2a). These results show that programmed cell death occurs specifically in female flies, reducing the number of mAL neurons.

We predicted the cell death programme was probably operating in the daughter cells rather than in the neuroblast itself. This is because *Df(3L)H99* homozygous neurons could be produced and labelled by heat-shock-induced activation of recombination at the late larval to

early pupal stage even in females (mosaics with labelled mAL neurons, 8/92 or 8.7%), which normally reveal no sign of labelling at this stage (Fig. 1q, r). Thus, in wild-type females, the cell death mechanism eliminates late-born neural progeny, whereas in neural progeny homozygous for *Df(3L)H99*, these cells survive and differentiate into neurons.

When we inhibited cell death at the embryonic stage by heat-shock-induced recombination of the *Df(3L)H99*-harbouring chromosome, the number of mAL neurons was increased in adult females. Such neurons had the contralateral forked arborization typical of females (Fig. 2d, circle). Notably, the ipsilateral projections never seen in normal females were formed in those female mAL neurons protected by the *Df(3L)H99* deletion (Fig. 2d, arrow). A single neuron with bilateral projections typical of males was stained in a female mAL cluster homozygous for *Df(3L)H99* (Fig. 2e, arrow). The contralateral projection of the neuron showed neither the typical female nor the typical male arborization pattern. This indicates that sex-specific cell death is not the sole mechanism for establishing sexually dimorphic neuronal projections, as the pattern of *H99*-protected neurons in the suboesophageal ganglion was not necessarily male-like.

The *NP21* strain has a P-element insertion in the second intron of the *fru* gene. Across the entire male brain, >80% of Gal4-expressing cells expressed Fru (Supplementary Fig. S2a–f). As illustrated in Fig. 3a–c, all of the mAL neurons expressing mCD8–GFP in males were Fru-positive. *NP21* homozygous males and *NP21/fru^{sat}* trans-heterozygous males showed similar abnormalities: loss of the male-specific muscle of Lawrence¹⁵ (Supplementary Fig. S2g–j), no expression of Fru (Supplementary Fig. S2k–n) and no detectable expression of intact male-specific *fru* transcripts by polymerase chain reaction with reverse transcription (RT–PCR) analysis (data not shown). We also observed markedly reduced courtship behaviour towards males as well as females (Supplementary Table S1). These results show that *NP21* is a loss-of-function allele of *fru*.

We next examined whether mutations in *fru* could affect the sexually dimorphic development of the mAL cluster (Fig. 3d). *NP21/fru^{sat}* males had only about 5 mAL neurons (Fig. 3e), a number comparable with that in females. The males of two other *fru* mutants also showed a significant reduction in the number of mAL neurons (Supplementary Fig. S2o, p). The similar reduction in the number of mAL cells in *fru* mutant males was confirmed by cell labelling using the MARCM method (Supplementary Fig. S2q–s). We performed single-progeny labelling in *NP21/NP21* flies, and counted about 5 mAL neurons in male flies (Fig. 3f).

In *NP21* homozygous males, no neurons with ipsilateral projection were observed. Instead, all mAL neurons in *NP21* homozygous males had contralateral, bifurcating projections reminiscent of their female counterparts (Fig. 3f). For males of other heteroallelic *fru* mutants, most mAL neurons had ‘female-like’ unilateral projections, but there were also neurons that had the extended ipsilateral projection typical of males (Fig. 3g and Supplementary Fig. S2r, s). It should be emphasized that the forked projections typical of females were formed in male *fru* mutants. To determine whether the Fru protein is required cell-autonomously in mAL neurons for male-typical development, *NP21/NP21* homozygous cells were generated in male flies heterozygous for *NP21/+*. mAL neurons with the *NP21/NP21* genotype had forked projections typical of females (Fig. 3h), indicating that cell-autonomous *fru⁺* function is necessary for male-type development of mAL neurons.

According to the hypothesis that Fru suppresses cell death specifically in males (Supplementary Fig. S3), Fru expression in female mAL neurons should inhibit the death of mAL neurons in females. Our experiments show that this is indeed the case (Fig. 3i–k): *tra¹* homozygous females express Fru in mAL neurons with male-like bilateral projections, which also have male-like arborizations in the suboesophageal ganglion. Thus, the *tra¹* mutation not only blocks cell death but also masculinizes the neuron branching pattern,

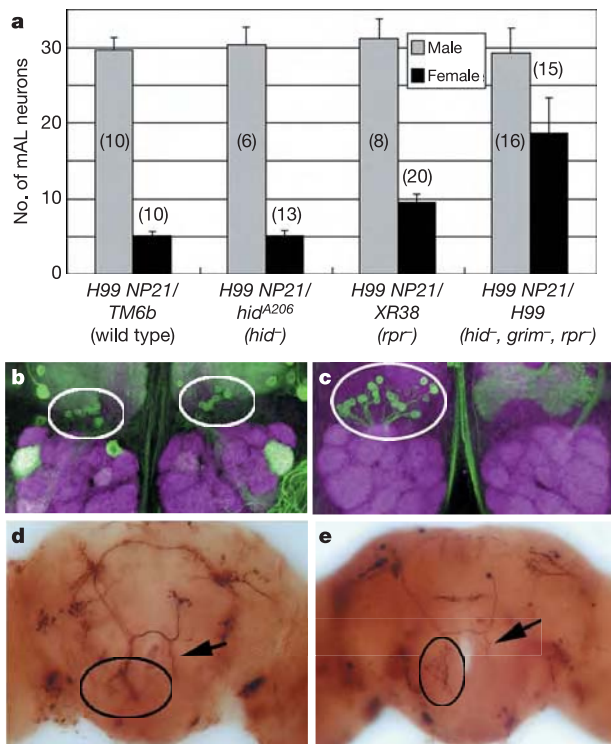


Figure 2 | Involvement of programmed cell death in the formation of mAL sexual dimorphism. **a**, The number of mAL neurons in males and females of control flies (*UAS-mCD8–GFP; Df(3L)H99 FRT 2A NP21/TM6B*), and three cell-death gene mutants, *hid^{A206}/Df(3L)H99* (*hid⁻* mutant), *Df(3L)XR38/Df(3L)H99* (*rpr⁻* mutant) and *Df(3L)H99/Df(3L)H99* (*hid⁻, grim⁻, rpr⁻* mutant) in the *NP21* heterozygous background. Numbers in parentheses indicate the number of brain hemispheres examined. Values are means ± s.d. **b**, **c**, Examples of brains stained with anti-GFP (green) and anti-nc82 (magenta) antibodies from adult females with genotypes *UAS-mCD8–GFP/+; Df(3L)H99 FRT 2A NP21/Df(3L)XR38* (**b**) and *yhs-flp; UAS-mCD8–GFP/+; Df(3L)H99 FRT 2A NP21/tub-Gal80 FRT 2A* (a mosaic) (**c**). **d**, **e**, Examples of mAL neurons in female brains stained with an anti-GFP antibody and visualized with HRP/DAB. Brains were dissected from adults (genotype *yhs-flp; UAS-mCD8–GFP/+; Df(3L)H99 FRT 2A NP21/tub-Gal80 FRT 2A*) in which recombination was induced at embryonic stage (**d**) or 4–5 days after egg collection (**e**) by heat-shock at 37 °C for 1 h. The mAL neurons in females retain a normal contralateral projection pattern (circle in **d**) but show ectopic ipsilateral projections (arrow in **d**). A single labelled neuron showed abnormally dispersed dendritic branching (circle in **e**) and ectopic ipsilateral projection (arrow in **e**) when cell death was inhibited.

presumably by inducing *Fru* expression in female mAL neurons. Such *tra*¹ mutant females vigorously courted partner females (Supplementary Table S1).

Here we have demonstrated that *Fru* expression is essential for the establishment of a 'male-typical' neural network. Our observations support the model that those neurons programmed to extend male-typical projections are eliminated in females by cell death (Fig. 1s, bottom). *Fru* inhibits cell death in male neurons at the pupal stage, allowing neurons to develop male-typical projections (Fig. 1s, top). A similar mechanism would operate in the formation of sexually dimorphic neurons in the optic lobe (Supplementary Fig. S4). *Fru* could thus be considered to function as a male–female switch in the CNS neural circuit.

The male-typical projections of mAL neurons in the suboesophageal ganglion have terminals that probably function as input sites. The suboesophageal ganglion is known to include the association centre for gustation¹⁶, and the central projections of both tarsal and labellar gustatory neurons terminate in the suboesophageal ganglion^{17–20}. Contact (gustatory) pheromones are also known to be involved in partner choice in *Drosophila* mating^{21–23}. Thus, an intriguing possibility is that the male-typical projections of mAL neurons in males act as the input sites of pheromonal information conveyed by gustatory neurons. The enhanced homosexual preference observed in *fru* mutant males might be due to a failure in integrating the pheromonal information crucial for partner recognition in mating behaviour, as a consequence of the loss of ipsilateral

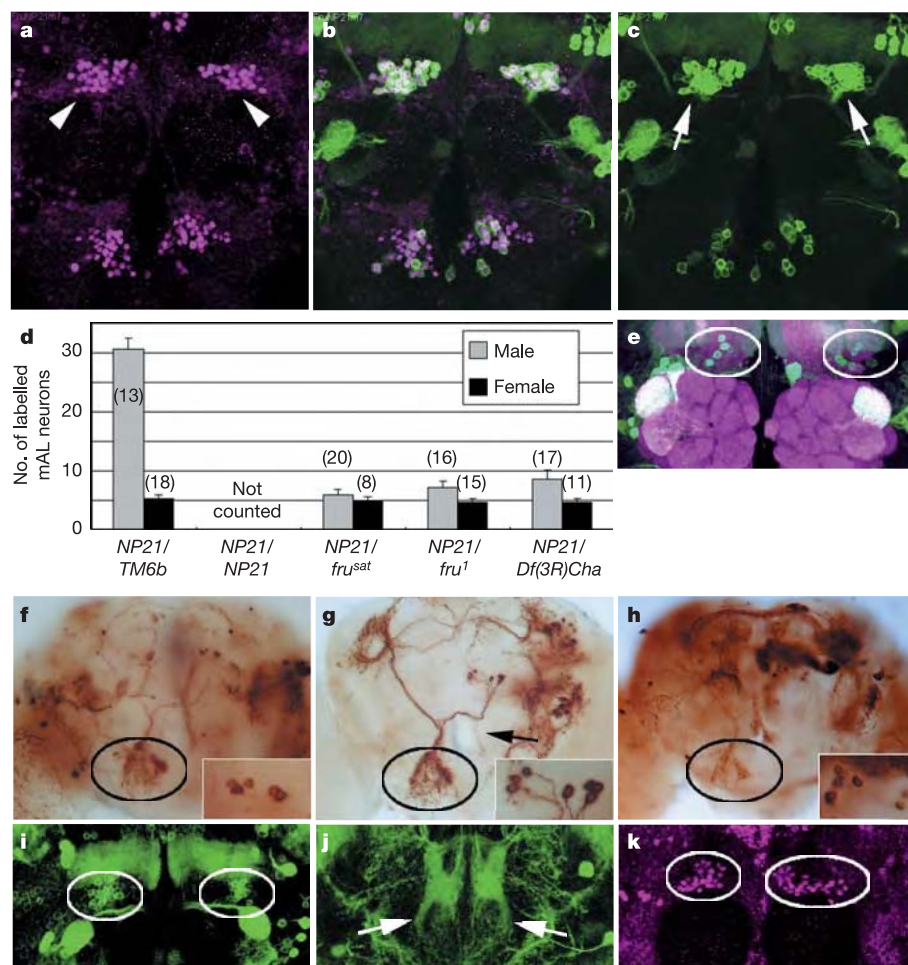


Figure 3 | Involvement of *fru* in the formation of mAL sexual dimorphism.

a–c, Male brain (*FRT G13 UAS-mCD8-GFP*; *NP21/TM6B*) doubly stained with anti-*Fru* (**a**, magenta) and anti-GFP (**c**, green) antibodies. The merged image is shown in **b**. Each of the 30 GFP-positive (and therefore Gal4-positive) mAL cells in (**c**) expresses *Fru* (arrowheads in **a**). **d**, The number of mAL neurons in male and female adults in different *fru* mutants, compared with the *NP21* heterozygous control (*FRT G13 UAS-mCD8-GFP*; *NP21/TM6B*). Numbers in parentheses indicate the number of brain hemispheres examined. Values are means $\pm$ s.d. We could not reliably determine the number of mAL neurons in *NP21* homozygotes, in which a large number of cells around the mAL cluster expressed mCD8-GFP, hampering the identification of authentic mAL neurons. **e**, An adult male brain (genotype *FRT G13 UAS-mCD8-GFP/+*; *NP21/fru*^{sat}) stained with anti-GFP (green) and anti-nc82 (magenta) antibodies. There were 5 mAL neurons in the male, comparable with the number in females. **f, g**, Examples of MARCM labelling of mAL neurons in the brains of *fru* mutant males stained with an anti-GFP antibody and visualized with HRP/DAB. Genotypes are *y hs-flp/Y*; *FRT G13 UAS-mCD8-*

GFP/FRT G13 tub-Gal80; *NP21/NP21* (**f**) and *y hs-flp/Y*; *FRT G13 UAS-mCD8-GFP/FRT G13 tub-Gal80*; *NP21/fru*^{sat} (**g**). In *NP21* homozygous males, all mAL neurons look like their female counterparts: male-specific ipsilateral projections are lost and they have female-like dendritic fields (circle in **f**). Similar changes in male mAL neurons were observed in *NP21/fru*^{sat} flies (circle in **g**), but some neurons retained male-typical projections (arrow in **g**). **h**, mAL neurons homozygous for *NP21* were generated in male brains of *NP21/+* background (genotype *y hs-flp/Y*; *UAS-mCD8-GFP/+*; *FRT 82 tub-Gal80/FRT 82 NP21*) using the MARCM method (heat-shock at 37 °C for 1 h was applied at the embryonic stage), stained with an anti-GFP antibody and visualized with HRP/DAB. A female-like pattern of arborization (circle) was formed cell-autonomously in the mAL cluster. Insets in **f–h** show magnified views of the region containing somata. **i–k**, A female brain (genotype *UAS-mCD8-GFP/+*; *tra*¹/*tra*¹ *NP21*) was stained with anti-GFP (green, **i, j**) or anti-*Fru* (magenta, **k**) antibodies. Female mAL neurons (circles in **i**) showed the male-like projection pattern in the suboesophageal ganglion (arrows in **j**) and also expressed *Fru* (circle in **k**).

projections from the mAL neurons. The study of Fru-expressing neurons and their circuits may therefore help to identify fundamental principles about the construction of sex-specific structures and functions in the CNS.

METHODS

Fly strains and clonal analysis. Flies were reared on cornmeal-yeast medium at 25 °C under constant illumination. Somatic clones were produced using the MARCM method¹¹. Eggs and/or embryos were collected within 24 h of egg laying, and after appropriate ageing were heat-shocked at 37 °C for 60 min.

CNS dissection, immunohistochemistry and imaging. Adult brains were dissected within 48 h of eclosion (emergence). Fixation and immunohistochemical staining were carried out as described²⁴, using antibodies at the following dilutions: rabbit polyclonal anti-GFP (1:1,000; Molecular Probes), mouse monoclonal nc82 (1:200; gift from A. Hofbauer), mouse monoclonal anti-GFP (1:200; Roche), rat monoclonal anti-HA (1:1,000; Roche), Cy3-conjugated goat anti-rabbit IgG (1:1,000), Cy2-conjugated goat anti-mouse IgG (1:200), Cy2-conjugated goat anti-rabbit IgG (1:1,000), Cy3-conjugated goat anti-mouse IgG (1:1,000), Cy5-conjugated goat anti-rabbit IgG (1:1,000), horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (1:2,000) (Jackson ImmunoResearch) and Alexa 647-conjugated goat anti-rat IgG (1:1,000; Molecular Probes). Two types of rabbit anti-Fru antibodies, Male-2 and MO5, were used at 1:500. Male-2 recognizes the male-specific N terminus, and MO5 recognizes the sequence common to all Fru isoforms. Both gave identical staining patterns in the CNS. Stacks of optical sections at 1 µm were obtained with an Olympus FV300 confocal microscope using FLUOVIEW software, and processed with Adobe Photoshop.

Analysis of courtship behaviour. The flies to be tested were reared individually in vials in a 12 h:12 h light:dark cycle, and aged for 4–7 days after eclosion. For courtship assays, a fly of each genotype was placed in a small chamber with a Canton-S virgin female or male at 4–7 days after eclosion, and their behaviour was videorecorded for 10 min. Courtship index (CI) was calculated as the percentage of time spent courting, including following, tapping, wing extension and attempted copulation.

Received 8 August; accepted 12 September 2005.

- Ito, H. *et al.* Sexual orientation in *Drosophila* is altered by the *satori* mutation in the sex-determination gene *fruitless* that encodes a zinc finger protein with a BTB domain. *Proc. Natl Acad. Sci. USA* **93**, 9687–9692 (1996).
- Ryner, L. C. *et al.* Control of male sexual behaviour and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87**, 1079–1089 (1996).
- Taylor, B. J. & Knittel, L. M. Sex-specific differentiation of a male-specific abdominal muscle, the Muscle of Lawrence, is abnormal in hydroxyurea-treated and in *fruitless* male flies. *Development* **121**, 3079–3088 (1995).
- Usui-Aoki, K. *et al.* Formation of the male-specific muscle in female *Drosophila* by ectopic *fruitless* expression. *Nature Cell Biol.* **2**, 500–506 (2000).
- Goodwin, S. F. *et al.* Aberrant splicing and altered spatial expression patterns in *fruitless* mutants of *Drosophila melanogaster*. *Genetics* **154**, 725–745 (2000).
- Nilsson, E. *et al.* *fruitless* is in the regulatory pathway by which ectopic *mini-white* and *transformer* induce bisexual courtship in *Drosophila*. *J. Neurogenet.* **13**, 213–232 (2000).
- Lee, G. *et al.* Spatial, temporal, and sexually dimorphic expression patterns of the *fruitless* gene in the *Drosophila* central nervous system. *J. Neurobiol.* **43**, 404–426 (2000).
- Demir, E. & Dickson, B. J. *fruitless* splicing specifies male courtship behaviour in *Drosophila*. *Cell* **121**, 785–794 (2005).
- Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behaviour. *Cell* **121**, 795–807 (2005).
- Manoli, D. S. *et al.* Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* **436**, 395–400 (2005).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
- White, K. *et al.* Genetic control of programmed cell death in *Drosophila*. *Science* **264**, 677–683 (1994).
- Peterson, C., Carney, G. E., Taylor, B. J. & White, K. *reaper* is required for neuroblast apoptosis during *Drosophila* development. *Development* **129**, 1467–1476 (2002).
- Grether, M. E., Abrams, J. M., Agapite, J., White, K. & Steller, H. The *head involution defective* gene of *Drosophila melanogaster* functions in programmed cell death. *Genes Dev.* **9**, 1694–1708 (1995).
- Lawrence, P. A. & Johnston, P. The muscle pattern of a segment of *Drosophila* may be determined by neurons and not by contributing myoblasts. *Cell* **45**, 505–513 (1986).
- Stocker, R. F. The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell Tissue Res.* **275**, 3–26 (1994).
- Thorne, N., Chromey, C., Bray, S. & Amrein, H. Taste perception and coding in *Drosophila*. *Curr. Biol.* **14**, 1065–1079 (2004).
- Wang, Z., Singhvi, A., Kong, P. & Scott, K. Taste representations in the *Drosophila* brain. *Cell* **117**, 981–991 (2004).
- Lienhard, M. C. & Stocker, R. F. Sensory projection patterns of supernumerary legs and arista in *D. melanogaster*. *J. Exp. Zool.* **244**, 187–201 (1987).
- Murphey, R. K., Possidente, D., Pollack, G. & Merritt, D. J. Modality-specific axonal projections in the CNS of the flies *Phormia* and *Drosophila*. *J. Comp. Neurol.* **290**, 185–200 (1989).
- Hall, J. C. The mating of a fly. *Science* **264**, 1702–1714 (1994).
- Greenspan, R. J. & Ferveur, F. J. Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**, 205–232 (2000).
- Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
- Awasaki, T. & Kimura, K. *pox-neuro* is required for development of chemosensory bristles in *Drosophila*. *J. Neurobiol.* **32**, 707–721 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Nilsson for critical reading of the manuscript, J. Maeyama and E. Matsuda for their participation in the early stages of this work, and M. Haga for technical assistance. We also thank K. Ito for the NP21 strain, A. Gould for the *Df(3L)H99 FRT A2* strain, K. White for the *Df(3L)XR38* strain, H. Steller for the *hid^{A206}* strain, A. Hofbauer for the monoclonal nc82 antibody, T. Awasaki, the Bloomington and Kyoto Stock Centers for *Drosophila* strains, and H. Aonuma for allowing us access to a confocal microscope. This work was supported in part by Special Cooperation Funds for Promoting Science and Technology from MEXT (Ministry of Education, Culture, Sports, Science and Technology in Japan) to K.-I.K. and D.Y.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.-I.K. (kimura@iwa.hokkyodai.ac.jp).

LETTERS

Suppression of Polycomb group proteins by JNK signalling induces transdetermination in *Drosophila* imaginal discs

Nara Lee¹, Cédric Maurange^{1,†}, Leonie Ringrose¹ & Renato Paro¹

During the regeneration of *Drosophila* imaginal discs, cellular identities can switch fate in a process known as transdetermination¹. For leg-to-wing transdetermination, the underlying mechanism involves morphogens such as Wingless that, when activated outside their normal context, induce ectopic expression of the wing-specific selector gene *vestigial*^{2,3}. Polycomb group (PcG) proteins maintain cellular fates by controlling the expression patterns of homeotic genes and other developmental regulators⁴. Here we report that transdetermination events are coupled to PcG regulation. We show that the frequency of transdetermination is enhanced in PcG mutant flies. Downregulation of PcG function, as monitored by the reactivation of a silent PcG-regulated reporter gene, is observed in transdetermined cells. This downregulation is directly controlled by the Jun amino-terminal kinase (JNK) signalling pathway, which is activated in cells undergoing regeneration. Accordingly, transdetermination frequency is reduced in a JNK mutant background. This regulatory interaction also occurs in mammalian cells, indicating that the role of this signalling cascade in remodelling cellular fates may be conserved.

Imaginal discs are clusters of cells that form the precursors of adult cuticular structures. On fragmentation and cultivation, a disc structure regenerates and forms the appendage, such as a leg or a wing, for which it was initially determined. Transdetermination events occur at sites of regeneration and on the ectopic expression of morphogens. Misregulation of PcG function causes homeotic transformations that are often phenocopied in transdetermination events. To determine whether PcG proteins are involved in transdetermination events, we induced leg-to-wing fate changes by ectopically overexpressing *wingless* (*wg*) and visualizing the transdetermined tissue by staining for ectopic expression of *vestigial* (*vg*) coupled to a *lacZ* reporter gene (Fig. 1a). In this assay, wild-type discs showed a transdetermination rate of 2.3% (2/87), whereas *Polycomb* (*Pc*) and *Sex combs on midleg* (*Scm*) heterozygous mutant discs showed an increased transdetermination rate of 17.7% (11/62) and 11.0% (10/91), respectively (Fig. 1a).

The PcG proteins function through *cis*-regulatory elements called PcG response elements (PREs), which enable them to bind and to maintain the state of transcriptional silencing over many cell divisions. PcG proteins operate in two key evolutionarily conserved chromatin complexes^{5–8}, and reduced expression of these complexes, as found in PcG mutants, results in the derepression of PRE-controlled genes. To determine whether PcG silencing is modulated in regenerating tissue, we used the FLW-1 line, which contains a *lacZ* reporter gene under the control of the *Fab7* PRE⁹. Prothoracic leg discs silent for *lacZ* expression were fragmented and transplanted into the abdomen of host flies. We fed flies with

5-bromodeoxyuridine (BrdU) to mark the regenerated tissue (the blastema). In uncut discs, there was little proliferation and expression of *lacZ* was undetectable (Fig. 1b). On fragmentation, however, *lacZ* was expressed in the blastema (Fig. 1c). To confirm that this derepression was due to a reduction in PcG silencing and not simply to massive proliferation at the wound site, we used the line LW-1 (ref. 9); this line lacks the *Fab7* PRE and is normally silent, but it can be activated by induction of GAL4 (ref. 9). Neither uncut nor cut leg discs of the LW-1 line showed expression of *lacZ* after transplantation (Supplementary Fig. 1a, b).

To show that transdetermination takes place only in cells with downregulated PcG function, we stained fragmented leg discs of the FLW-1 line for *lacZ* expression and for *Vg* to visualize the transdetermination to wing fate (Fig. 1d). We consistently observed that the *Vg* staining lay within the *lacZ* expression domain, suggesting that PcG genes are downregulated in the blastema, enabling PRE-silenced genes to be reactivated according to new morphogenetic cues.

To investigate direct targets of PcG regulation that, when reactivated, might contribute to transdetermination, we tested the PREs predicted at the *wg* and *vg* genes and found that both are controlled by PcG proteins (Supplementary Fig. 2). The fact that both the transgenic *vg-lacZ* reporter construct (which lacks the PRE) and the endogenous *vg* gene were upregulated in the blastema suggests that PcG proteins may affect *vg* expression both indirectly (for example, through *wg*) and directly by means of the *vg* PRE.

JNK signalling in *Drosophila* is crucial for wound healing^{10–12} and is implicated in many different developmental processes, such as dorsal and thorax closure¹³ (Fig. 2a). *hemipterous* encodes the JNK kinase (JNKK)¹⁴ that activates the *Drosophila* JNK Basket¹⁵. Products of *DJun*¹⁶ and *kayak* (the *Drosophila* homologue of Fos¹⁷) form the AP-1 transcription factor. A downstream target of JNK signalling is *puckered* (*puc*), which encodes a phosphatase that selectively inactivates Basket and thus functions in a negative feedback loop¹⁸. The expression of *puc* thus mirrors JNK activity. Because wound healing takes place after fragmentation, we reasoned that activation of the JNK pathway might be causing the downregulation of PcG proteins in the blastema. We used the *puc*^{E69} line, which carries a P(*lacZ*) insertion at the *puc* locus, to monitor JNK activity. During the third-instar larval stage *puc* is not expressed and thus JNK signalling was not activated in leg discs (Fig. 2b). As expected, however, *puc* was expressed on fragmentation in all cells at the annealing cut edge (Fig. 2c).

To check whether cells that have activated the JNK pathway also show transdetermination, we transplanted fragmented leg discs of flies carrying the *puc-lacZ* reporter and *vg*^{BE}-*Gal4*; *UAS-GFP* constructs.

¹Centre for Molecular Biology Heidelberg (ZMBH), University of Heidelberg, 69120 Heidelberg, Germany. [†]Present address: The National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

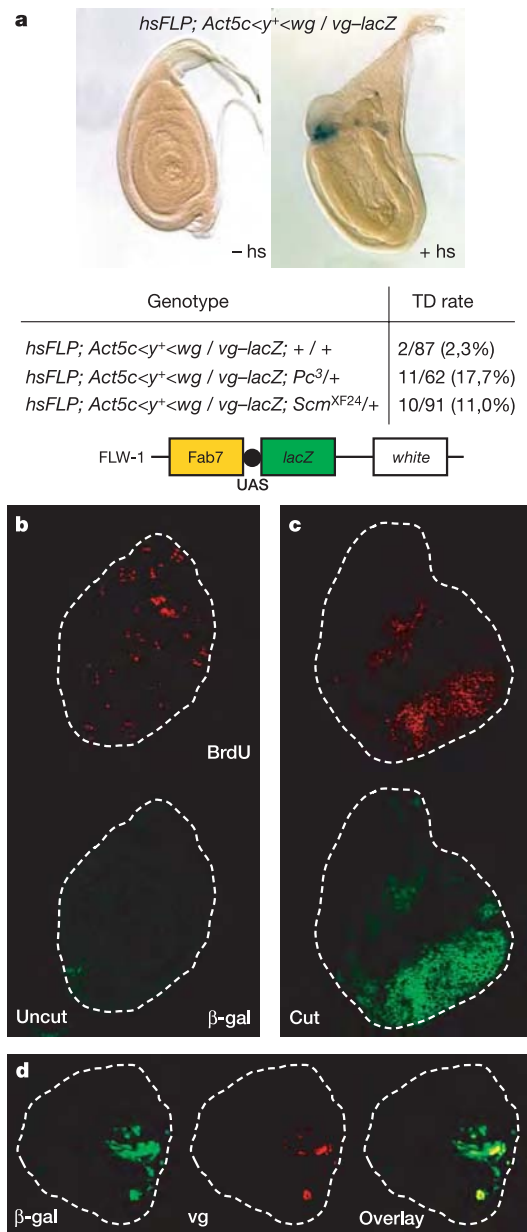


Figure 1 | PcG genes control transdetermination events. **a**, The *wg*-induced transdetermination rate is enhanced in PcG heterozygous mutants. Transdetermination to wing fate was scored by the ectopic expression of *vg^{BE}-lacZ*. The transdetermination rate was 2.3% in wild-type flies, and 17.7% and 11.0% in *Pc* and *Scm* heterozygous mutant flies, respectively. *hs*, heatshock. **b**, **c**, PcG genes are downregulated in the blastema. The structure of the reporter construct of FLW-1 flies is shown on top. Prothoracic leg discs of FLW-1 flies were stained for BrdU and β -galactosidase (β -gal). In **b**, a transplanted uncut disc is shown; in **c**, a fragmented disc is shown after *in vivo* culture. β -Gal staining is observed only in regenerated cells. **d**, *Vg* is expressed only in the region of PcG downregulation. Co-staining of fragmented leg discs of the FLW-1 line showed that in all cases tested *Vg* expression lies within the region stained by β -gal. The genotype in **b–d** is *FLW1; hsGal4/+*. Experiments were done at 25 °C under non-induced expression of GAL4.

In these flies, cells that adopted a wing fate were identified by their expression of green fluorescent protein (GFP; Fig. 2e). Two days after fragmentation, weak residual *puc-lacZ* staining was still visible in the central region of the disc (Fig. 2d, brackets). *puc-lacZ* staining is known to decline rapidly after wound healing is completed¹¹. We found that stronger staining was visible along the cut site, probably owing to ongoing wound healing. On comparison of *puc-lacZ*

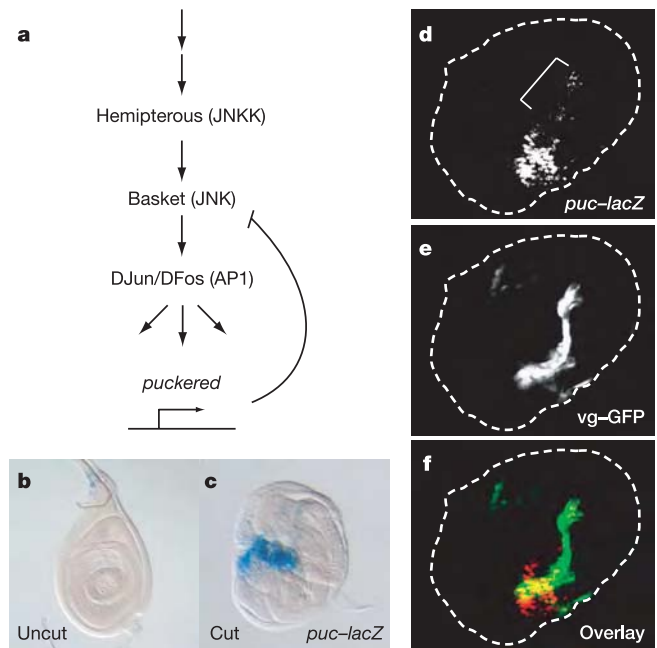


Figure 2 | JNK signalling is activated at the fragmentation site. **a**, The *Drosophila* JNK signalling pathway. **b**, *puc* is not expressed in third-instar prothoracic leg discs. **c**, After fragmentation, *puc* is expressed at the wound site as a result of healing where the cut edges anneal (1 d after *in vivo* culture). **d–f**, *Puc*- and *Vg*-induced GFP staining substantially overlap. **d**, After 2 d of *in vivo* culture *puc* is still expressed. Weak residual staining is seen in the centre (bracket); stronger staining is seen along the cut site. **e**, The GFP fluorescence of *vg^{BE}-Gal4; UAS-GFP* flies indicates transdetermined cells. **f**, Merged image.

staining and GFP fluorescence, JNK-active cells showed a substantial overlap with transdetermined cells (Fig. 2f); thus, we conclude that JNK signalling is activated in cells that undergo transdetermination.

JNK signalling affects the transcription of numerous genes, including those encoding chromatin regulating factors¹⁹. We therefore examined whether JNK signalling can downregulate the PcG proteins required for transdetermination. We overexpressed a constitutively active form of *hep* in *UAS-hep^{act}; hsGal4* flies by a heat-shock pulse²⁰. Activating the JNK pathway caused a downregulation of some PcG genes, such as *Pc*, *ph-p* and *E(Pc)* (Fig. 3a, b). We observed no downregulation of these genes in wild-type larvae before and after heat shock, indicating that this was not an unspecific heat-shock response (data not shown). We examined the expression of two genes of the Trithorax group (*ash1* and *brm*) that function antagonistically to PcG proteins, but found no upregulation on JNK induction (data not shown).

To show further that JNK has a specific effect on PcG proteins, we carried out the analogous experiment in mammalian cells. The JNK pathway can be activated in mouse embryonic fibroblasts by exposing the cells to ultraviolet light²¹. We examined the expression of *Mph2* (mouse *polyhomeotic2*) because this mammalian PcG gene is expressed in these cells. The expression of *Mph2* was decreased on JNK induction (Fig. 3c), but after treatment with a specific JNK inhibitor²² it was partially restored. In addition, to show that the downregulation of PcG genes is directly controlled by AP-1, we carried out chromatin immunoprecipitation using antibodies against Fos on chromatin from *UAS-hep^{act}; hsGal4* and *kay¹* mutant flies. We observed enrichment of Fos on the promoter region of *ph-p*, but no enrichment in chromatin from flies lacking Fos (Fig. 3d, e). This finding suggests that AP-1 binds directly to this region to regulate negatively the transcription of *ph-p*.

If activation of JNK signalling in the blastema indeed leads to a downregulation of PcG genes, then impairment of the JNK pathway

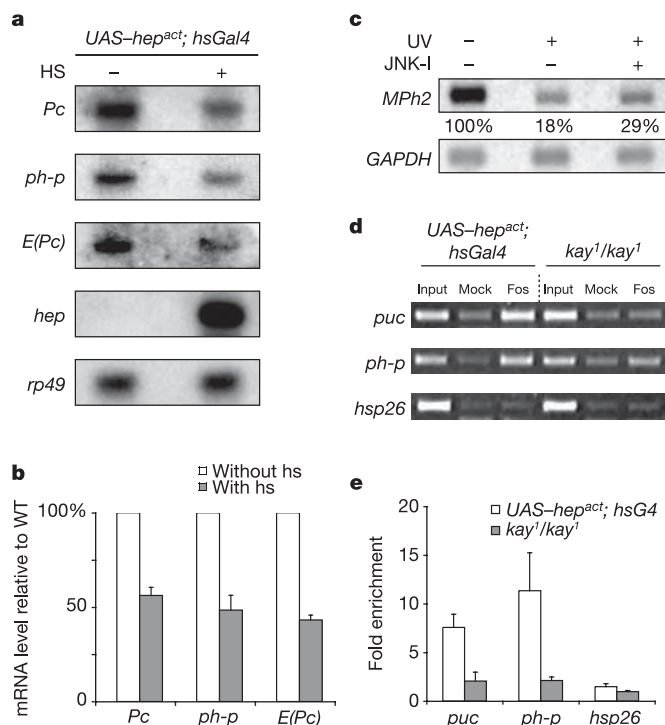


Figure 3 | Activation of JNK leads to downregulation of PcG genes.

a, Northern blot of RNA isolated from whole third-instar larvae showing downregulation of *Pc*, *ph-p* and *E(Pc)* after activation of JNK. The blot was probed for *hep* to verify that heat shock was effective. *rp49* was used as a loading control. **b**, Quantification of the downregulation of PcG genes. Data are from three independent experiments; error bars indicate the s.d. **c**, Northern blot of mouse embryonic fibroblasts shows a decrease in *Mph2* RNA after activation of JNK by ultraviolet treatment. Treatment with a JNK inhibitor (D-JNK-I) partially restores the expression of *Mph2*. *GAPDH* was used as a loading control. **d**, Crosslinked chromatin was immunoprecipitated with antibodies against Fos. PCR was used to detect promoter fragments of *puc* as a positive control, *hsp26* as negative control and *ph-p*. **e**, Quantification of the chromatin immunoprecipitation. Data are from three independent experiments; error bars indicate the s.d.

should result in reduced efficiency of transdetermination. We compared the transdetermination behaviour of wild-type discs with that of discs bearing mutations in the JNKK *hep*. The transdetermination events were classified into three categories: large regions, small regions, and no regions of transdetermination. In wild-type discs only large regions were detected (34/34; Fig. 4). In males hemizygous for *hep*¹ (a weak hypomorphic allele¹⁴), most transplanted leg discs had large transdetermined regions (20/37); however, a substantial proportion showed only small regions of transdetermination (14/37) and a few showed no transdetermination event (3/37). In flies heterozygous for *hep*⁷⁵ (a null allele which is hemizygous lethal), most discs showed no or only small regions of transdetermination (21/27), and large regions were rarely seen (3/27). The morphology of the regenerated discs seemed unaffected in these mutants, indicating that the decline of transdetermination efficiency was not due to inefficient wound healing.

We have shown that PcG genes are downregulated by JNK signalling. Because many developmental regulators need to be switched, the role of PcG downregulation may be to render the cells susceptible to a change in cell identity by shifting the chromatin to a reprogrammable state. Transdetermination has been ascribed to the action of ectopic morphogens, which induce cells to activate incorrect gene cascades^{23,24}. Without doubt, *wg* and *decapentaplegic* signalling must be crucially involved in this process, because transdetermination does not result from any random cut but occurs preferentially when cuts are made through particular regions of the

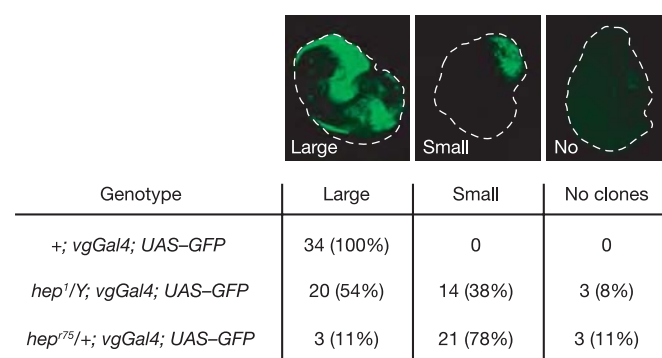


Figure 4 | Impaired JNK signalling reduces the transdetermination ratio.

Transdetermination events were classified into large and small regions. In wild-type discs, only large transdetermined regions were observed (34/34). In *hep*¹/*Y* hemizygous males, most discs showed large regions (20/37); however, a substantial proportion showed small regions (14/37) and a few discs show no detectable transdetermination (3/37). *hep*⁷⁵/+ flies showed a few discs with either large regions (3/27) or no transdetermination (3/27), but mostly discs with small regions (21/27).

disc called 'weak points', which are regions of high morphogen²⁴. Inappropriate or overextreme downregulation of the PcG system by JNK in sensitive cells of the weak points thus may create such aberrant local patterns. Indeed, our data indicate that at least the two patterning genes, *wg* and *vg*, may be direct targets of the PcG (Supplementary Fig. 2). Notably, hyperactive Wnt signalling can also induce a switch in lineage commitment in mammals²⁵, implying that signalling pathways are a potent inducer of cell fate changes in many organisms.

Another study has shown that regenerating and transdetermining cells in the blastema have a distinct cell-cycle profile in contrast to the surrounding normal disc cells³. It was proposed that this change in cell-cycle regulation is a prerequisite for the change in cell fate. Indeed, PcG targets include genes involved in cell-cycle regulation^{26,27}, suggesting that this initial step is part of the complete reprogramming cascade required for the regenerating cells to achieve multipotency. Downregulation of PcG silencing by JNK seems to be a fundamental, evolutionarily conserved mechanism of cell fate change and thus may also have implications for studies of stem cell plasticity and tissue remodelling.

METHODS

Fly stocks. Flies were raised on standard media at 25 °C. *w*¹¹¹⁸ flies were used as wild-type controls. *Act5c* < *y*⁺ < *wg* flies were obtained from K. Basler; *FLW-1* and *LW-1* (*U15* 1,1) have been described⁹; *hep*¹ and *hep*⁷⁵ flies¹⁴ were a gift from S. Noselli; *kay*¹ flies were provided by the Tübingen Stock Collection; *puc*^{E69} (ref. 18) and *UAS-hep*^{act} (ref. 20) were obtained from D. Bohmann; *vg*^{BE}-*lacZ* and *vg*^{BE}-*Gal4* from S. Cohen; *yw* *hsFLP* and *UAS-GFP* flies from Bloomington. **Transdetermination and in vivo culture.** We carried out *wg*-induced transdetermination as described². In brief, *hsFLP*; *Act5c* < *y*⁺ < *wg*/*vg*^{BE}-*lacZ* flies were heat-shocked during second instar for 1 h at 37 °C. Subsequently, prothoracic leg discs of third-instar larvae were stained with X-Gal as described⁹. Fragmentation of prothoracic leg discs was done as described²⁸. Using a tungsten needle, leg discs of mid third-instar larvae were cut in Ringer's solution to generate fragments comprising the posterior three-quarters of the disc. These were transplanted into the abdomen of etherized female hosts, which had been collected shortly after eclosion and kept for 1 d at 18 °C without food. After transplantation, the host flies were kept at 25 °C for 3 d unless otherwise stated, in the presence of males, on standard food supplemented with yeast paste. For BrdU staining, host flies were fed BrdU, which was mixed into the food at a final concentration of 200 µg ml⁻¹. Regenerated leg discs were recovered from host abdomens in PBS and fixed for immunohistochemistry.

Chromatin immunoprecipitation. Embryos of *kay*¹/*kay*¹ and *UAS-hep*^{act}/+; *hsGal4*/+ genotype (aged 10–12 h) were collected and heat-shocked for 1 h at 37 °C and, after 1 h at 25 °C, chromatin was isolated using standard protocols. Immunoprecipitation was done using antibodies against Fos (a gift from D. Bohmann) and PCR was performed with primers that amplify the promoter

region of *puc* as a known AP-1 target, *hsp26* as a negative control, and *ph-p*. We used the following primer sequences: *puc*, 5'-ggtttgagcccgagataa-3' and 5'-actgaagacttgcggtgaa-3'; *hsp26*, 5'-ttaataagaggaaaccag-3' and 5'-aaaaataaaactaactt-3'; *ph-p*, 5'-tgcgtcattttgttc-3' and 5'-tggcgcggcatcacttt-3'.

Northern blot analysis. The northern blotting procedure is described in the Supplementary Information.

Received 21 June; accepted 27 July 2005.

- Hadorn, E. In *The Genetics and Biology of Drosophila* Vol. 2c (eds Ashburner, M. & Wright, T.) 556–617 (Academic, New York, 1978).
- Maves, L. & Schubiger, G. Wingless induces transdetermination in developing *Drosophila* imaginal discs. *Development* **121**, 1263–1272 (1995).
- Sustar, A. & Schubiger, G. A transient cell cycle shift in *Drosophila* imaginal disc cells precedes multipotency. *Cell* **120**, 383–393 (2005).
- Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
- Otte, A. P. & Kwaks, T. H. Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? *Curr. Opin. Genet. Dev.* **13**, 448–454 (2003).
- Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37–46 (1999).
- Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
- Muller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
- Cavalli, G. & Paro, R. The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505–518 (1998).
- Ramet, M., Lanot, R., Zachary, D. & Manfrulli, P. JNK signalling pathway is required for efficient wound healing in *Drosophila*. *Dev. Biol.* **241**, 145–156 (2002).
- Galko, M. J. & Krasnow, M. A. Cellular and genetic analysis of wound healing in *Drosophila* larvae. *PLoS Biol.* **2**, E239 (2004).
- Bosch, M., Serras, F., Martin-Blanco, E. & Baguna, J. JNK signalling pathway required for wound healing in regenerating *Drosophila* wing imaginal discs. *Dev. Biol.* **280**, 73–86 (2005).
- Zeitlinger, J. & Bohmann, D. Thorax closure in *Drosophila*: involvement of Fos and the JNK pathway. *Development* **126**, 3947–3956 (1999).
- Glise, B., Bourbon, H. & Noselli, S. hemipterous encodes a novel *Drosophila* MAP kinase kinase, required for epithelial cell sheet movement. *Cell* **83**, 451–461 (1995).
- Riesgo-Escovar, J. R., Jenni, M., Fritz, A. & Hafen, E. The *Drosophila* Jun-N-terminal kinase is required for cell morphogenesis but not for DJun-dependent cell fate specification in the eye. *Genes Dev.* **10**, 2759–2768 (1996).
- Perkins, K., Dailey, G. & Tjian, R. Novel Jun- and Fos-related proteins in *Drosophila* are functionally homologous to enhancer factor AP-1. *EMBO J.* **7**, 4265–4273 (1988).
- Riesgo-Escovar, J. R. & Hafen, E. Common and distinct roles of DFos and DJun during *Drosophila* development. *Science* **278**, 669–672 (1997).
- Martin-Blanco, E. *et al.* *puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* **12**, 557–570 (1998).
- Jasper, H. *et al.* The genomic response of the *Drosophila* embryo to JNK signalling. *Dev. Cell* **1**, 579–586 (2001).
- Weber, U., Paricio, N. & Mlodzik, M. Jun mediates Frizzled-induced R3/R4 cell fate distinction and planar polarity determination in the *Drosophila* eye. *Development* **127**, 3619–3629 (2000).
- Shaulian, E. *et al.* The mammalian UV response: c-Jun induction is required for exit from p53-imposed growth arrest. *Cell* **103**, 897–907 (2000).
- Bonny, C., Oberson, A., Negri, S., Sauser, C. & Schorderet, D. F. Cell-permeable peptide inhibitors of JNK: novel blockers of β -cell death. *Diabetes* **50**, 77–82 (2001).
- Maves, L. & Schubiger, G. A molecular basis for transdetermination in *Drosophila* imaginal discs: interactions between wingless and decapentaplegic signalling. *Development* **125**, 115–124 (1998).
- Maves, L. & Schubiger, G. Transdetermination in *Drosophila* imaginal discs: a model for understanding pluripotency and selector gene maintenance. *Curr. Opin. Genet. Dev.* **13**, 472–479 (2003).
- Okubo, T. & Hogan, B. L. Hyperactive Wnt signalling changes the developmental potential of embryonic lung endoderm. *J. Biol.* **3**, 11 (2004).
- Ringrose, L., Rehmsmeier, M., Dura, J. M. & Paro, R. Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev. Cell* **5**, 759–771 (2003).
- Valk-Lingbeek, M. E., Bruggeman, S. W. & van Lohuizen, M. Stem cells and cancer; the Polycomb connection. *Cell* **118**, 409–418 (2004).
- Gibson, M. C. & Schubiger, G. Hedgehog is required for activation of engrailed during regeneration of fragmented *Drosophila* imaginal discs. *Development* **126**, 1591–1599 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Gehring for teaching us the transplantation technique; G. Schubiger for advice; D. Bohmann for the anti-Fos antibody and fly lines; S. B. Carroll for the anti-Vg antibody; K. Basler, S. Noselli, S. Cohen, J. Grosshans and the Tübingen Fly Stock Collection for fly lines; S. Dlugosz and F. Zimmermann for technical support with injection needles; O. Pein and P. Angel for mouse embryonic fibroblast cells and advice; C. Beisel, S. Schönfelder and M. Tariq for comments and discussions. R.P. was funded by the Deutsche Forschungsgemeinschaft.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.P. (paro@zmbh.uni-heidelberg.de).

LETTERS

A resetting signal between *Drosophila* pacemakers synchronizes morning and evening activity

Dan Stoleru^{1*}, Ying Peng^{1*}, Pipat Nawathean¹ & Michael Rosbash¹

The biochemical machinery that underlies circadian rhythms is conserved among animal species and drives self-sustained molecular oscillations and functions, even within individual asynchronous tissue-culture cells^{1–3}. Yet the rhythm-generating neural centres of higher eukaryotes are usually composed of interconnected cellular networks, which contribute to robustness and synchrony as well as other complex features of rhythmic behaviour^{4–7}. In mammals, little is known about how individual brain oscillators are organized to orchestrate a complex behavioural pattern. *Drosophila* is arguably more advanced from this point of view: we and others have recently shown that a group of adult brain clock neurons expresses the neuropeptide PDF⁸ and controls morning activity (small LN_v cells; M-cells), whereas another group of clock neurons controls evening activity (CRY⁺, PDF[–] cells; E-cells)^{6,9}. We have generated transgenic mosaic animals with different circadian periods in morning and evening cells. Here we show, by behavioural and molecular assays, that the six canonical groups of clock neurons¹⁰ are organized into two separate neuronal circuits. One has no apparent effect on locomotor rhythmicity in darkness, but within the second circuit the molecular and behavioural timing of the evening cells is determined by morning-cell properties. This is due to a daily resetting signal from the morning to the evening cells, which run at their genetically programmed pace between consecutive signals. This neural circuit and oscillator-coupling mechanism ensures a proper relationship between the timing of morning and evening locomotor activity.

Overexpression of the TIM kinase SHAGGY (SGG; *Drosophila* GSK3) shortens the period by 3–4 h (ref. 11). We first drove SGG expression in all clock cells by crossing *tim-GAL4* (ref. 12) with flies carrying an EP element inserted at the *SGG* locus (*EP1576*, referred to as *UAS-SGG*)^{11,13}. The locomotor activity rhythm of *tim-GAL4/UAS-SGG* (*timSGG*) flies in constant darkness (DD) confirmed previous results¹¹ (Fig. 1a and Supplementary Fig. S1b), in that the period was about 3 h shorter than that of control flies ($\tau_{timSGG} = 20.6 \pm 0.3$ h; $\tau_{UAS-SGG} = 23.7 \pm 0.2$ h; τ is the average locomotor period; $\pm$ s.e.m.).

We next expressed SGG only in LN_v cells by constructing a *Pdf-GAL4/UAS-SGG* genotype. The *Pdf-GAL4* driver is well characterized⁸ and drives gene expression only in two clock-cell groups: the PDF⁺ small LN_v (s-LN_v) cells (that is, M-cells)⁹ and the PDF⁺ large LN_v (l-LN_v) cells. The driver is inactive in the CRY⁺PDF[–] evening cells⁶. *Pdf-GAL4/UAS-SGG* (*PdfSGG*) flies also manifested a short period ($\tau_{PdfSGG} = 21.7 \pm 0.2$ h; Fig. 1c). The period shortening was less than that of *timSGG* flies, probably because of weaker expression from *Pdf-GAL4* driver in LN_v cells⁶. (SGG expression from an even weaker driver, *cry₁₃-GAL4*, did not affect behavioural period (Supplementary Fig. S1a).)

A closer inspection of the behavioural actograms (Supplementary

Fig. S1b) revealed that the period of evening activity is significantly shorter in *PdfSGG* flies (with a daily advance of about 2 h; Supplementary Fig. S1b). This indicates that the pace of E-cells was accelerated, although the period manipulation was restricted to M-cells. An advanced evening peak, without an increase in E-cell SGG expression, indicates that the faster M oscillator might be setting the E-cell pace. We therefore proposed that the PDF⁺ cells influence molecular circadian events within E-cells.

To investigate this possibility, we estimated the molecular period (cycle duration) of each clock-cell group in these different genotypes: *UAS-SGG* (control), *timSGG* and *PdfSGG*. Fly brains were analysed by *in situ* hybridization for *tim* RNA expression pattern after 4 days in DD, so that a barely detectable daily advance by 2–3 h would result in an aggregate advance of 8–12 h on DD4 (fourth day of DD). Indeed, SGG overexpression in all clock neurons (*timSGG*) markedly shifted the interval of high *tim* mRNA expression on DD4 by about 12 h, from between CT10 and CT18 to before CT6 (compare Fig. 1b with Fig. 1a, right panels). (CT is the circadian time within a constant-darkness experiment; CT0 is the hour of the last lights-on event.) All neurons expressing clock genes showed a similar temporal pattern, consistent with the expected SGG-induced period shortening in all clock cells, and with a deterministic relationship between the molecular period and the locomotor activity period.

However, the *PdfSGG tim* RNA profiles were strikingly different and unexpected (Fig. 1c and Supplementary Figs S6 and S7). Whereas the s-LN_v cells showed a roughly 8 h advance in DD4, expected from a period shortening of 2 h per day, the l-LN_v cells showed no appreciable change from those in control flies; that is, their molecular programme was apparently unaffected by SGG overexpression within these cells (Fig. 1c and Supplementary Figs S6 and S7). Also surprising were the DN1 and DN3 profiles, which showed a roughly 8 h advance, as were the LN_d cells, which were advanced by about 6 h relative to those in control flies (Fig. 1a, c; Supplementary Figs S2 and S6). As *PdfSGG* flies do not overexpress SGG in these three cell groups, their molecular programmes behave in a non-cell-autonomous manner. Because the E-cells are included within these groups⁶ and because the s-LN_v cells (the M-cells)⁹ are the only cells with a cell-autonomous programme that matches the behavioural period of the flies, the M-cells apparently determine the clock pace of these other neuronal groups, including the E-cells.

The l-LN_v cells and DN2 cells emerged as the only clock-gene-expressing neurons that evaded control of the M-cells and maintained a wild-type-like phase of *tim* RNA cycling in *PdfSGG* flies. Because DN2 cells are genotypically wild type in these flies, we infer that they oscillate with cell-autonomous properties and are the best candidates for determining the non-cell-autonomous wild-type-like characteristics of the l-LN_v cells. As a consequence there are at least two parallel clock-cell circuits in the *Drosophila* brain in constant

¹Howard Hughes Medical Institute and National Center for Behavioral Genomics, Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA.

*These authors contributed equally to this work.

darkness: the M–E circuit controls locomotor activity rhythms and is driven by the M-cells (s-LN_v cells), whereas the DN2–l-LN_v circuit has as yet unknown functions and is driven by the DN2 cells.

To verify and extend these concepts, we generated a genotype in which the E-cells should run faster than M-cells. By adding the previously described *Pdf-GAL80* repressor construct to the *tim-GAL4;UAS-SGG* background⁶, SGG was expected to be overexpressed in all clock neurons with the exception of PDF-expressing cells. As these include the M-cells (s-LN_v cells), they should run more slowly (24 h) than the E-cells (about 21 h). A ‘faster takes all’ rule predicts that the short-period E-cells will dominate over the normal 24 h M-cells in this genotype and generate a behavioural rhythm of about 21 h. Alternatively, dominant M-cells will give rise to a behavioural period of 24 h despite the faster endogenous oscillator in the E-cells.

Consistent with a dominant M-cell model was the observation that *timSGG/PdfGAL80* flies had an almost wild-type period in DD, $\tau = 23.8 \pm 0.2$ h (Fig. 2a, left). The molecular analysis was also consistent, as the s-LN_v cells manifested a wild-type-like programme:

tim mRNA peaked between CT12 and CT20 on DD4 (Fig. 2a, right). Despite SGG overexpression, the LN_d cells, DN1 cells and DN3 cells had a similar and wild-type-like pattern of *tim* expression (Fig. 2a, and Supplementary Figs S8 and S9). As described above, this indicates that all three cell groups behave non-autonomously and are probably driven by the s-LN_v cells. This result is supported by the anatomical pattern of s-LN_v neuronal processes, which project towards the brain regions populated by LN_d, DN1 and DN3 cells^{12,14}. DN2 cells were again the only SGG-overexpressing cells in which the phase of *tim* RNA oscillation corresponded to the predicted accelerated pace (Fig. 2b). The l-LN_v cells, despite lacking SGG overexpression (because of the *PdfGAL80* transgene), also showed a comparable advance of *tim* expression (Fig. 2b). These *timSGG/PdfGAL80* results confirm that the s-LN_v cells determine the phase of LN_d, DN1 and DN3 cells and that an independent cellular network includes the DN2 and l-LN_v cells. Because the behavioural period was wild-type-like and paralleled the molecular clock within the s-LN_v cells, the results confirm that these M-cells assign the circadian period in the absence of light cues.

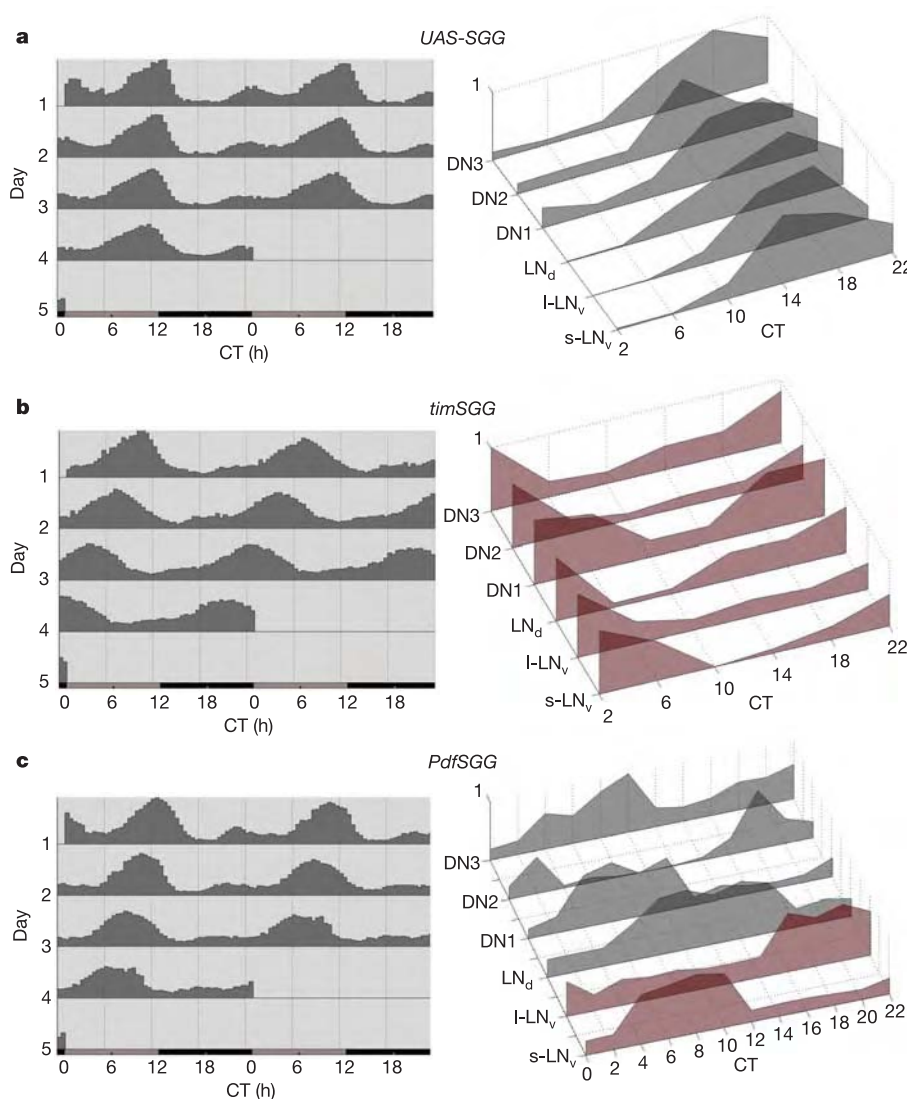


Figure 1 | The M-cell clock influences the timing of events related to E-cells. Characterization of behavioural and molecular phenomena in flies in which different neuronal groups run molecular circadian programmes with different periods. Left panels: double-plotted actograms of locomotor behaviour, during the first four days in constant darkness (DD1–4). Right panels: plots representing the temporal changes in *tim* RNA levels during

DD4. The cell groups in which SGG was overexpressed are plotted in red. **a**, *UAS-SGG* flies ($n = 110$; $\tau = 23.6$ h); **b**, *timSGG* flies ($n = 100$; $\tau = 20.8$ h); **c**, *PdfSGG* flies ($n = 125$; $\tau = 21.7$ h). For a more detailed cell-group-specific analysis of the *tim* signal oscillations, see the main text and Supplementary Figs S2–S11.

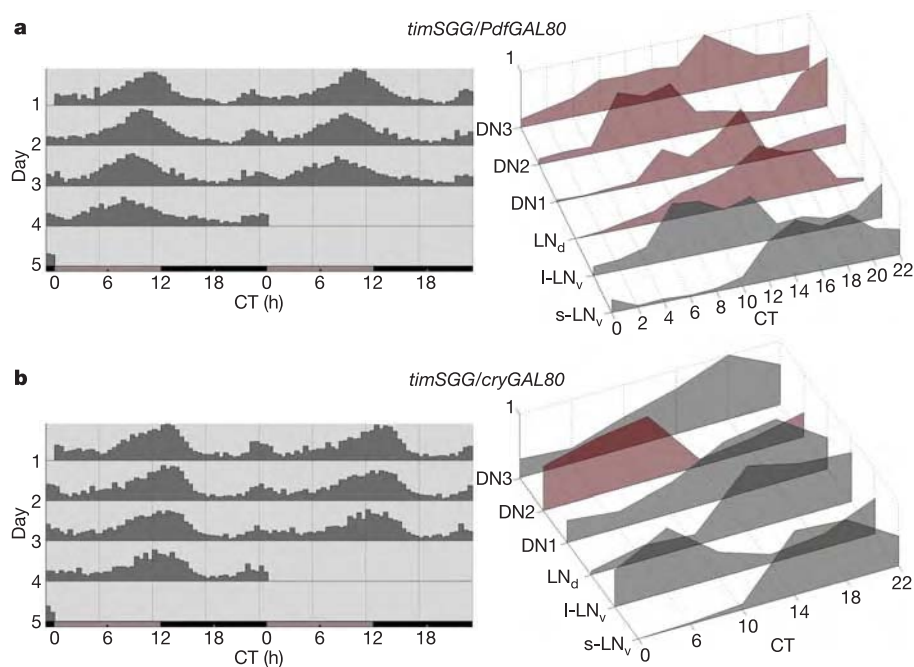


Figure 2 | The M-cell clock dictates the period and phase of E-cell molecular oscillation, and DN2 cells control the phase of I-LN_v cells. **a**, *timSGG/PdfGAL80* flies ($n = 32$; $\tau = 23.7$ h); **b**, *timSGG/cryGAL80* flies ($n = 32$; $\tau = 23.8$ h). The panels are structured as in Fig. 1.

To confirm the lack of a contribution of DN2/I-LN_v to the E-M network function and to locomotor rhythms, we also examined the *timSGG/cryGAL80* genotype. It is similar to the *timSGG/PdfGAL80* genotype described above, except that SGG overexpression is repressed in a wider group of cells. These include most if not all of the E-cells and I-LN_v cells as well as the M-cells⁶. As DN2 cells are

only clock cells in which *cry* promoter-driven expression was not detected⁶, we expected that the faster clock in *timSGG/cryGAL80* would be limited to CRY⁺ cells, including the apparently cell-autonomous DN2 cells.

Indeed, *tim* hybridization *in situ* confirmed that the period of DN2 rhythm was shortened by about 2–3 h per day (Fig. 2b and

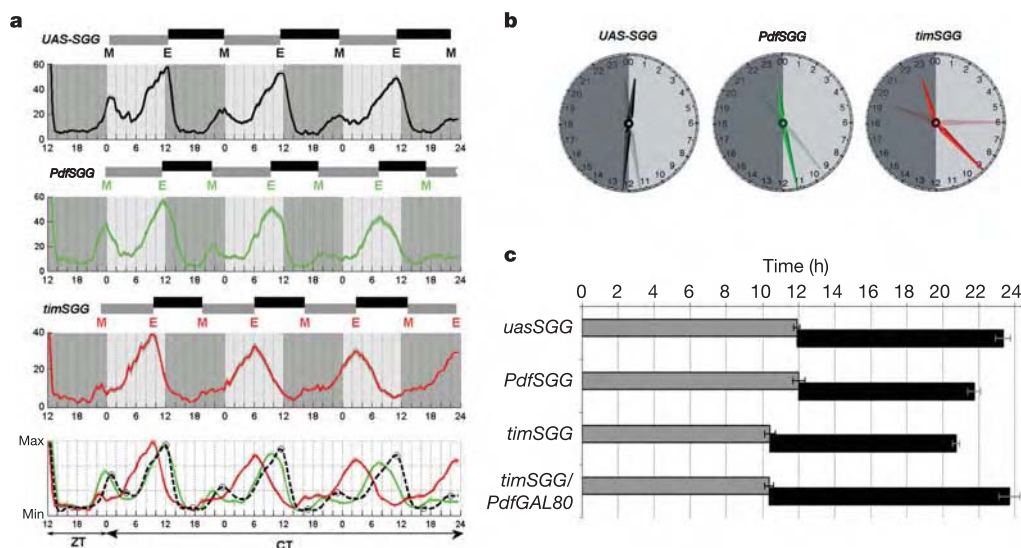


Figure 3 | The duration of subjective day is correlated with the genotype of E-cells, whereas the period is correlated with the genotype of M-cells. Phase comparison of locomotor activity for three genotypes: *UAS-SGG* (black), *PdfSGG* (green) and *timSGG* (red). **a**, The top three panels show a time-based analysis of DD locomotor activity peaks for each genotype: the starting time point coincides with the last light-off event, that is, the last night of LD (ZT12–24) and the following three days in DD are shown. The timing is indicated by different background shades of grey: light grey for CT0–12 (corresponding to light interval in LD) and dark grey for CT12–24 (corresponding to dark interval in LD). The M and E peaks are indicated above each panel. The bottom panel shows a three-way genotype

comparison of the peak phases. The automatically identified peaks are marked with open circles (control), asterisks (*timSGG*) and diamonds (*PdfSGG*). The x axis shows ZT and CT time in hours, and the y axis shows mean activity. **b**, The actual timing of morning and evening peaks during the first 48 h after lights-off. The numbers on the 24 h dials are as shown in **a**. The short arm indicates the time of the morning peak, the long one the time of evening activity. The full-coloured indicators represent the timing of events during the first cycle of activity, and the shaded ones refer to the second cycle after lights-off. **c**, Comparison of average duration of subjective day (M–E, in grey), during DD1–4. Black bars show the length of subjective night (E–M). Error bars represent s.d.

Supplementary Figs S10, S11). The $l-LN_v$ neurons were shifted to about the same extent, which is consistent with the notion that they behave non-cell-autonomously and follow the pace of the DN2 clock programme. All other clock cells maintained a pattern similar to that of control flies (Fig. 2b). Because *timSGG/cryGAL80* flies had a normal behavioural period ($\tau = 23.9 \pm 0.1$ h; Fig. 2a), these results confirm that $l-LN_v$ and DN2 cells do not contribute detectably to locomotor activity rhythms. This conclusion is in agreement with previous results showing that wild-type flies have persistent DD behavioural rhythms, despite protein oscillation idiosyncrasies of the $l-LN_v$ and DN2 cells^{15,16}.

How does the M-cell ($s-LN_v$) clock determine the period of E-cells (LN_d cells/DN cells)? Although our previous work indicated possible oscillator coupling⁶ and a direct effect of LN_v on the transcriptional oscillations of other clock cells⁵, it was difficult to envision how the M-cells could override the intrinsic molecular timing of the E-cells. We therefore considered a second possibility, namely that the E-cells maintain an unaltered intrinsic clock programme but receive a daily resetting signal from the M-cells. This model predicts that the timing of the evening activity within every cycle (between two consecutive mornings) reflects the status of the endogenous clock of E-cells, whereas the overall period exhibited by the evening peaks reflects the pace of the M-cell resetting signal.

To examine this possibility, we assayed the different transgenic strains for their average evening activity phase within a cycle, by

using the leading morning peak as a reference and then measuring the average time until the subsequent evening peak (M–E interval = subjective day; Fig. 3). The overall DD period correlated with the genotype of M-cells as expected, but the length of the subjective day (M–E interval) correlated only with the genotype of the E-cells. In control flies (*UAS-SGG*) with a period of about 24 h, the subjective day was roughly 12 h, similar to the duration of subjective day of *PdfSGG* (Fig. 3b, c). The latter strain features a wild-type-like E-oscillator but a fast, SGG-expressing M-oscillator and a period of about 22 h (Fig. 1a). In contrast, *timSGG* flies express SGG in both E-cells and M-cells, and both the average length of subjective day and the period (M–M) are reduced (M–E = 10.45 h; period more than 3 h shorter; Figs 1b and 3b, c). Taken together with the molecular data (Fig. 1, right panels; Supplementary Figs S2–S7), the results indicate that the E-cells run an autonomous clock programme whose starting (or ending) points are determined by daily resetting signals from the M-cells.

A DD unidirectional M → E resetting mechanism also predicts that a slower (24 h) M-cell clock and a faster E-cell clock will have a normal morning peak phase but an advanced evening peak phase. To test this prediction we compared the behavioural outputs of *timSGG/PdfGAL80* and *timSGG/cryGAL80* flies, which differ only in the genotypes of their E-cells (Fig. 4). Both strains have periods of about 24 h, but the former should give rise to a fast E-cell molecular programme, whereas the latter should have an E-clock of 24 h as a result of suppression of SGG expression.

Indeed, the evening phase of *timSGG/cryGAL80* is similar to that of control flies (Fig. 4b, middle), and it always occurs about 2.5 h later than that of *timSGG/PdfGAL80* (Fig. 4). The evening phase of *timSGG/PdfGAL80* is more similar to that of *timSGG* (Fig. 4b, right), although the latter genotype ($\tau = 20.6$ h) has a much shorter period than the former ($\tau = 23.8$ h). The length of subjective day of *timSGG/PdfGAL80* flies (M–E = 10.35 h; that is, similar to that of *timSGG*; Fig. 3c) further confirms that the evening phase within each cycle is a reflection of the endogenous E-cell rhythm, whereas the period of the cycle (M–M) correlates with the intrinsic M-cell clock.

These comparisons indicate that the circadian network is modulated by intercellular communication signals, which achieve and maintain circadian coherence—the proper relationship between morning and evening activity. The dominant M-clock determines the period of the entire system by providing a daily reset signal to the E-clock in darkness and is therefore a true cellular Zeitgeber. Because the M-cells can delay as well as advance E-cells, the resetting signal may be required for E-cell oscillations. The usual candidate for this signal is the M-cell-specific neuropeptide PDF. It contributes to the normal synchrony and/or rhythmicity in constant darkness^{5,7}, with a striking similarity to the mammalian neuropeptide VIP^{4,17}. Moreover, injecting PDF into the cockroach brain causes circadian phase delays¹⁸. Other principles and/or molecules may also be relevant to the M–E subnetwork, because E-cells can drive clockless M-cells to manifest cyclical behavioural outputs under 12 h light/12 h dark (LD) conditions⁶.

The $l-LN_v$ and DN2 cells are the two neuronal groups that escape the M-cell reset signal in DD. They constitute a second circadian subnetwork with no apparent effect on locomotor activity rhythms and no known function. The DN2 cells are among the few clock-gene-expressing brain cells in larvae¹² and are also the only clock cells that do not change their morphology after eclosion¹⁹. Larval DN2 cells are apparently devoid of CRY and manifest anti-phase oscillations of TIM and PER^{10,19}. It is therefore likely that both the DN2 cells and the $l-LN_v$ cells impart circadian regulation to unknown physiological functions relevant to both larvae and adults. More generally, we expect that the organizational principles of the two subnetworks described here will also be relevant to mammalian neuronal networks with important behavioural functions, for example the relationship between different oscillators in the SCN^{20,21}.

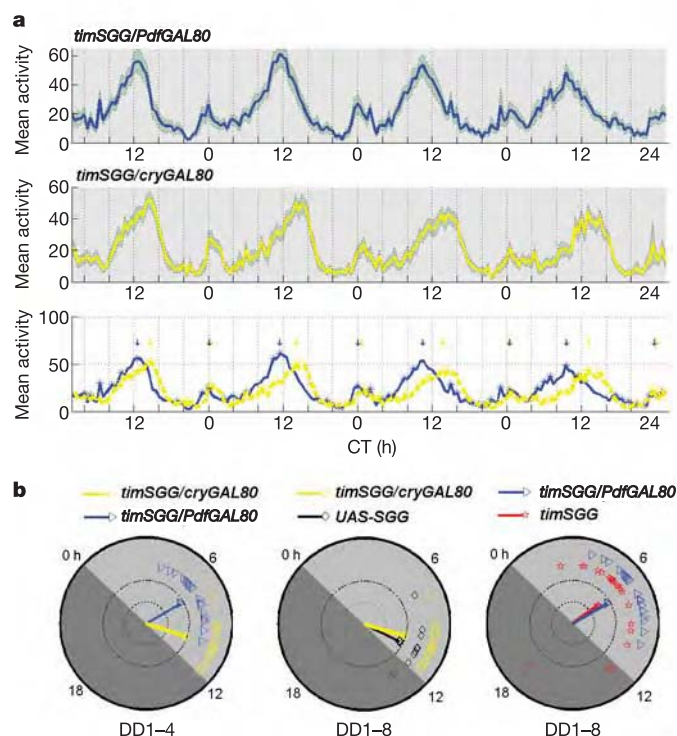


Figure 4 | The evening oscillator times its output according to its own intrinsic clock programme. **a**, Phase comparison of the DD locomotor activity of *timSGG/PdfGAL80* flies (blue) and *timSGG/cryGAL80* flies (yellow). In the bottom panel, the M and E peaks of each genotype are indicated by small arrows of corresponding colour. The graph shows counts of activity over time. **b**, Circular phase analysis of DD1–4 and DD1–8 locomotor activity²³. The vectors indicate the average phase of the evening peak for the whole group²³. The length of the vector is proportional to the statistical strength and coherence within the group²³. The evening locomotor activity of *timSGG/cryGAL80* flies has a similar phase to that of control flies (bivariate test; confidence = 99.9%), and the evening activity of *timSGG/PdfGAL80* is similar to that of *timSGG* (bivariate test; confidence = 100%).

METHODS

Fly strains. *Pdf-GAL4* (ref. 8), *tim-GAL4* (refs 6, 12), *cry₁₃-GAL4*, *Pdf-GAL80^{96A}* and *cry-GAL80^{2e3m}* (ref. 6) transgenic lines have all been characterized previously. The *EP1576* line was obtained from the Bloomington Stock Center (<http://fly.bio.indiana.edu/>). All molecular and behavioural experiments were performed at 25 °C.

Behavioural analysis. Flies were entrained for 5 days in LD conditions before being released into conditions of constant darkness (DD). Locomotor activities of individual male flies were monitored with Trikinetics *Drosophila* activity monitors. Analysis was performed with a signal-processing toolbox²² implemented in MATLAB (MathWorks). Autocorrelation and spectral analysis were used to assess rhythmicity and calculate period. Phase estimation was achieved by using circular statistics²³. For the calculation of M–E intervals in DD (Fig. 3b, c), we took the following steps: we applied a Butterworth filter to the activity time courses to smooth out random peaks (those oscillating with periods of less than 20 h)²². We then manually chose the first peak between two consecutive evening events (which were detected unambiguously by the software²²) and identified its corresponding phase as morning CT time. The time interval between every morning peak and the following evening peak was computed for the first four days in darkness, averaged and represented in Fig. 3c. The interval between two subsequent morning peaks was similarly calculated and represented. To reduce noise, more than 100 flies from each genotype were used. Alternative methods of estimating the phase of morning/evening activities (including the activity onset, centre of activity mass, increasing the number of individuals in a group to about 300 or varying the values of the smoothing filter) were also used and gave similar results (data not shown).

Imaging techniques. Fly brains were dissected and mounted as described previously⁵. *In situ* hybridization of *Pdf* and *tim* was also conducted as described²⁴. At least ten brains were examined for each time and genotype. The maximum projection images taken with a Leica laser scanning confocal microscope were used for quantifying *tim* RNA signals with ImageJ software (<http://rsb.info.nih.gov/ij/>). Values for individual cells were computed as the ratio between the pixel intensity of the cell and the average intensity of three background areas in neighbouring brain regions. Six to ten single-cell *tim* signals from at least four different hemispheres were averaged to obtain the mean values for each of the six individual cell groups (for example the DN1 group) for every time point, and standard deviations were calculated. Mean values within every given experiment were then normalized to the maximum value of the six (Figs 1a, b and 2c, right panels) or 12 (Figs 1c and 2a) time points. The detailed quantification results are shown in Supplementary Information, together with samples of brain pictures (Supplementary Figs S2–S11). The identification of PDF⁺ LN_d and DN cells was based on their specific location and anatomy (size and number). The particularly problematic distinction between DN1 cells and the two DN2 cells was based mainly on the position of the latter group in a slightly more posterior coronal section (layer).

Received 22 June; accepted 2 September 2005.

1. Welsh, D. K., Yoo, S. H., Liu, A. C., Takahashi, J. S. & Kay, S. A. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Curr. Biol.* **14**, 2289–2295 (2004).
2. Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).
3. Nagoshi, E. *et al.* Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* **119**, 693–705 (2004).
4. Hattar, A. *et al.* The VPAC₂ receptor is essential for circadian function in the mouse suprachiasmatic nuclei. *Cell* **109**, 497–508 (2002).
5. Peng, Y., Stoleru, D., Levine, J. D., Hall, J. C. & Rosbash, M. *Drosophila* free-running rhythms require intercellular communication. *PLoS Biol.* **1**, E13 (2003).
6. Stoleru, D., Peng, Y., Agosto, J. & Rosbash, M. Coupled oscillators control

morning and evening locomotor behaviour of *Drosophila*. *Nature* **431**, 862–868 (2004).

7. Lin, Y., Stormo, G. D. & Taghert, P. H. The neuropeptide pigment-dispersing factor coordinates pacemaker interactions in the *Drosophila* circadian system. *J. Neurosci.* **24**, 7951–7957 (2004).
8. Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C. & Taghert, P. H. A *pdf* neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioural circadian rhythms in *Drosophila*. *Cell* **99**, 791–802 (1999).
9. Grima, B., Chelot, E., Xia, R. & Rouyer, F. Morning and evening peaks of activity rely on different clock neurons of the *Drosophila* brain. *Nature* **431**, 869–873 (2004).
10. Kaneko, M., Helfrich-Forster, C. & Hall, J. C. Spatial and temporal expression of the *period* and *timeless* genes in the developing nervous system of *Drosophila*: newly identified pacemakers candidates and novel features of clock gene product cycling. *J. Neurosci.* **17**, 6745–6760 (1997).
11. Martinek, S., Inonog, S., Manoukian, A. S. & Young, M. W. A role for the segment polarity gene *shaggy*/GSK-3 in the *Drosophila* circadian clock. *Cell* **105**, 769–779 (2001).
12. Kaneko, M. & Hall, J. C. Neuroanatomy of cells expressing clock genes in *Drosophila*: Transgenic manipulation of the *period* and *timeless* genes to mark the perikarya of circadian pacemaker neurons and their projections. *J. Comp. Neurol.* **422**, 66–94 (2000).
13. Rorth, P. A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. *Proc. Natl Acad. Sci. USA* **93**, 12418–12422 (1996).
14. Helfrich-Forster, C. The period clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **92**, 612–616 (1995).
15. Veleri, S., Brandes, C., Helfrich-Forster, C., Hall, J. C. & Stanewsky, R. A self-sustaining, light-entrainable circadian oscillator in the *Drosophila* brain. *Curr. Biol.* **13**, 1758–1767 (2003).
16. Yang, Z. & Sehgal, A. Role of molecular oscillations in generating behavioural rhythms in *Drosophila*. *Neuron* **29**, 453–467 (2001).
17. Aton, S. J., Colwell, C. S., Hattar, A. J., Waschek, J. & Herzog, E. D. Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. *Nature Neurosci.* **8**, 476–483 (2005).
18. Petri, B. & Stengl, M. Pigment-dispersing hormone shifts the phase of the circadian pacemaker of the cockroach *Leucophaea maderae*. *J. Neurosci.* **17**, 4087–4093 (1997).
19. Klarsfeld, A. *et al.* Novel features of cryptochrome-mediated photoreception in the brain circadian clock of *Drosophila*. *J. Neurosci.* **24**, 1468–1477 (2004).
20. de la Iglesia, H. O., Cambras, T., Schwartz, W. J. & Diez-Noguera, A. Forced desynchronization of dual circadian oscillators within the rat suprachiasmatic nucleus. *Curr. Biol.* **14**, 796–800 (2004).
21. Jagota, A., de la Iglesia, H. O. & Schwartz, W. J. Morning and evening circadian oscillations in the suprachiasmatic nucleus *in vitro*. *Nature Neurosci.* **3**, 372–376 (2000).
22. Levine, J., Funes, P., Dowse, H. & Hall, J. Signal analysis of behavioural and molecular cycles. *BMC Neurosci.* **3**, 1 (2002).
23. Levine, J., Funes, P., Dowse, H. & Hall, J. Advanced analysis of a *cryptochrome* mutation's effects on the robustness and phase of molecular cycles in isolated peripheral tissues of *Drosophila*. *BMC Neurosci.* **3**, 5 (2002).
24. Zhao, J. *et al.* *Drosophila* clock can generate ectopic circadian clocks. *Cell* **113**, 755–766 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Hall, R. Allada, P. Emery and R. Stanewsky for critical comments on the manuscript; J. Agosto, J. Menet, E. Nagoshi and S. Kadener for discussions and advice; E. Dougherty for assistance in confocal microscopy; and H. Felton for administrative assistance. The work was supported in part by grants from the NIH to M.R.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.R. (rosbash@brandeis.edu).

Principal pathway coupling agonist binding to channel gating in nicotinic receptors

Won Yong Lee^{1,2} & Steven M. Sine¹

Synaptic receptors respond to neurotransmitters by opening an intrinsic ion channel in the final step in synaptic transmission. How binding of the neurotransmitter is conveyed over the long distance to the channel remains a central question in neurobiology. Here we delineate a principal pathway that links neurotransmitter binding to channel gating by using a structural model of the *Torpedo* acetylcholine receptor at 4-Å resolution¹, recordings of currents through single receptor channels and determinations of energetic coupling between pairs of residues. We show that a pair of invariant arginine and glutamate residues in each receptor α -subunit electrostatically links peripheral and inner β -sheets from the binding domain and positions them to engage with the channel. The key glutamate and flanking valine residues energetically couple to conserved proline and serine residues emerging from the top of the channel-forming α -helix, suggesting that this is the point at which the binding domain triggers opening of the channel. The series of interresidue couplings identified here constitutes a primary allosteric pathway that links neurotransmitter binding to channel gating.

Throughout the central and peripheral nervous systems, synaptic transmission depends on postsynaptic receptors to detect released

neurotransmitter and change the membrane potential. In rapidly activated neurotransmitter receptors, the agonist-binding site protrudes into the synaptic cleft but is far from the channel that forms the barrier to ions and gates their flow. A network of loops at the interface separating binding and channel domains has been shown to link agonist binding to channel gating^{2,3}; however, the atomic scale pathway that functionally links the two domains remains unknown.

Identifying the linkage pathway has been limited by the lack of an atomic structure of the coupling region. However, an improved electron density map of the *Torpedo* acetylcholine (ACh) receptor at 4-Å resolution has now resulted in an atomic scale model¹. Inspection of the model suggests that a series of molecular links in the α -subunit forms a pathway from the ACh-binding site to the channel gate (Fig. 1). The pathway begins with a structure known as the C-loop, which disengages from the binding site in the absence of agonist^{1,4,5}, but caps the site when agonist is bound^{4,6,7}. This capping motion may propagate to the interface separating binding and channel domains through rigid-body motion of part of the C-loop known as β -strand 10. At the end of β -strand 10, Arg 209 extends into the hydrophobic core of the subunit, where it juxtaposes Glu 45 in an electrostatic contact. Glu 45 and the flanking Val 46 branch from the

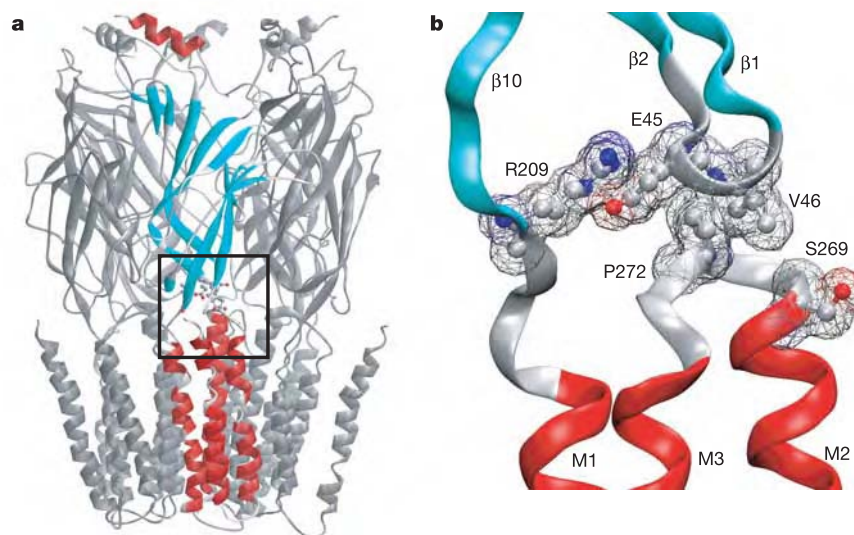


Figure 1 | Structural model of the *Torpedo* receptor ligand-binding and pore domains at 4 Å resolution (PDB 2BG9) and proposed principal pathway linking agonist binding to channel gating. **a, Panoramic image showing the pentameric structure; the secondary structures of one α -subunit are highlighted (α -helices, red; β -sheets, cyan; loops, grey). **b**, Close-up view of the principal pathway in the α -subunit; atoms of key residues are**

highlighted according to electrostatic charge (blue, positive; red, negative, grey, neutral). van der Waals surfaces are rendered as a wire mesh. The key residues branch from β -strand 10 of the C-loop, the β 1– β 2 linker and the linker spanning M2 and M3. The Cys-loop, β 8– β 9 linker and M4 helix have been removed for clarity. Images were generated with VMD²⁹.

¹Receptor Biology Laboratory, Department of Physiology and Biomedical Engineering, and ²Molecular Neuroscience Graduate Program, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA.

$\beta 1$ – $\beta 2$ hairpin and straddle Pro 272 in the linker that follows the channel-forming M2 transmembrane helix. Val 46 also fits into the cavity created by residues Ser 269 to Pro 272. Thus, the molecular continuum described here potentially links the binding site to the channel.

To determine whether this molecular continuum links binding to gating, we recorded single-channel currents through nicotinic receptors cloned from human skeletal muscle containing mutations of the key residues. These residues are identical in *Torpedo* and human receptors. We first tested the pair Arg 209 and Glu 45 by charge reversal (Fig. 2a). The charge-reversal mutation R209E abolishes expression of receptors on the cell surface, preventing

functional analysis, but the charge-neutralizing mutation R209Q allows expression and strongly impairs channel gating; channel openings are abbreviated and the intervening closings are prolonged (Fig. 2b). The charge-reversal mutation E45K shows a similarly impaired kinetic fingerprint, indicating that Glu 45 makes comparable contributions to channel gating and possibly interacts electrostatically with Arg 209.

To quantify changes in channel gating, we analysed single-channel open and closed dwell times to obtain rate constants in a mechanism in which ACh-binding and channel gating are separate, sequential events^{8,9} (Methods). For both R209Q and E45K, the rate constant for channel opening slows, whereas that for channel closing accelerates, resulting in a 45- to 120-fold decrease in the channel gating equilibrium constants (Table 1). Rate–equilibrium free-energy relationships (REFER) suggest that channel gating results from a sequence of domain shifts in which the binding domain attains the conformation of the open state first and the gating domain attains it last¹⁰. Thus, for residues at the interface separating binding and channel domains, we expect an intermediate domain shift with intermediate changes in rate as a function of equilibrium constant; indeed, REFER analysis of channel gating applied to wild-type, R209Q, E45K, and E45R receptors gives the expected intermediate shift, and a plot of gating rate against equilibrium constants yields a slope, ϕ , of 0.43 ± 0.03 ($\pm$ standard error).

To test further for interaction between Arg 209 and Glu 45, we generated pairs of charge-reversal mutations in individual α -subunits and recorded single-channel currents from the resulting receptors. The combination R209E and E45K prevents cell-surface expression, but the combination R209E and E45R yields functional surface receptors; thus, E45R rescues expression abolished by R209E alone. Moreover, the charge-reversal receptor responds to ACh and gates the channel with high efficiency (Fig. 2b). Kinetic analysis of single-channel dwell times yields a channel gating equilibrium constant of 27 for the wild-type receptor¹¹, whereas that for the charge-reversal mutant is 20 (Table 1). Thus, Arg 209 and Glu 45 are interchangeable and interdependent.

To quantify energetic coupling between Arg 209 and Glu 45, we generated receptors with the mutations R209Q, E45R and the corresponding double mutation, recorded ACh-evoked single-channel currents, and cast the resulting channel gating equilibrium constants as a mutant cycle¹² (Fig. 2c). Throughout, agonist binding was verified by the ability of ACh to reduce α -bungarotoxin binding to that of nonspecific binding (Supplementary Methods), and the channel gating equilibrium constant was measured because it directly registers communication between the binding site and the channel. When present individually, both R209Q and E45R impair channel gating, but combining the two mutations restores gating to near normal (Table 1). When the gating equilibrium constants are cast as a mutant cycle, the free energy of coupling is $-3.1 \pm 0.1 \text{ kcal mol}^{-1}$ (Fig. 2c). This magnitude of the coupling energy is similar to that measured for a salt bridge in the hydrophobic interior of the protein barnase ($-3.3 \text{ kcal mol}^{-1}$)¹³. Thus, Arg 209 and Glu 45 satisfy several criteria expected of a principal pathway linking binding to gating: they are located at the interface between binding and channel domains, are physically proximal, are found in all members of the Cys-loop receptor superfamily (Supplementary Table S3), and are functionally coupled in contributing to channel gating.

In other members of the Cys-loop receptor superfamily, residues at positions equivalent to Arg 209 contribute to agonist potency and efficacy^{14,15}, although pairwise interactions have remained elusive. The positions flanking Arg 209 may also contain a positively charged residue, such as Lys 215 of the $\beta 2$ subunit of the γ -amino butyric acid A (GABA_A) receptor, which precedes the position equivalent to Arg 209 and contributes to agonist potency¹⁶, and Arg 222 of the 5-HT_{3A} receptor, which follows the position equivalent to Arg 209 and contributes to agonist potency and efficacy¹⁷. Thus, positively

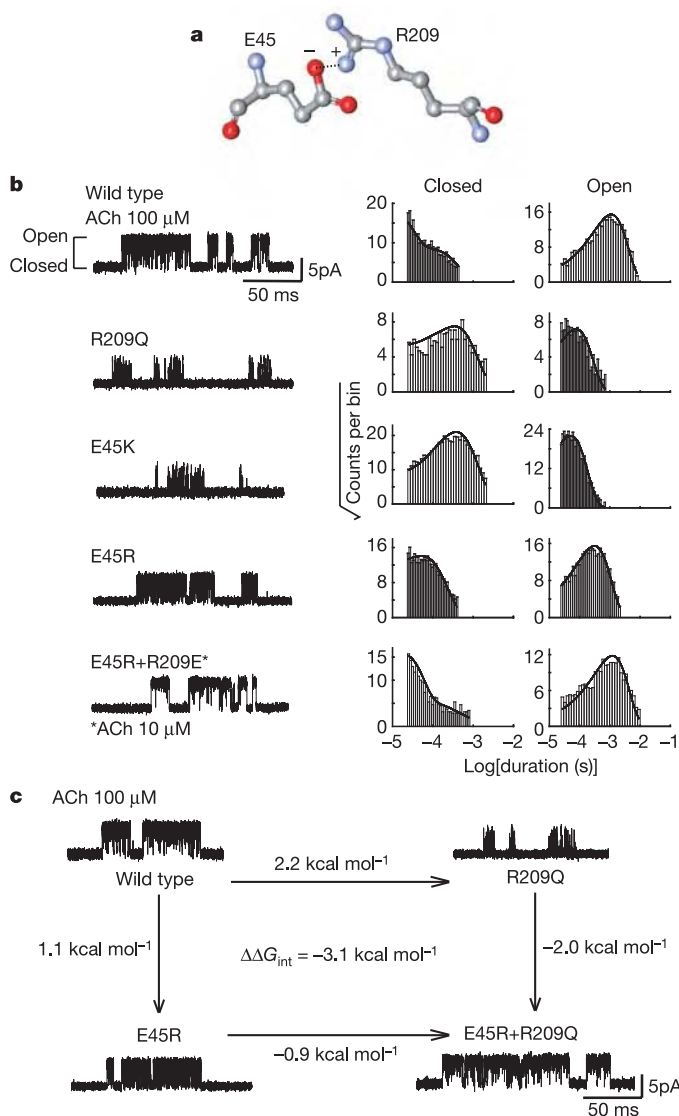


Figure 2 | Energetic coupling between Glu 45 and Arg 209 in the α -subunit. **a**, Glu 45 and Arg 209 are displayed using the coordinates from PDB 2BG9. **b**, Single-channel currents (10-kHz bandwidth) elicited by 100 μ M ACh through the indicated receptors expressed in BOSC 23 cells. The corresponding open and closed time histograms are shown, along with probability density functions (smooth curves) computed from the fitted binding and gating rate constants (see Methods, Table 1 and Supplementary Table 1). **c**, Mutant cycle analysis of the indicated receptors. Single-channel currents elicited by 100 μ M ACh are shown above each receptor type. Free energy changes for channel gating are shown for each limb of the cycle; the overall coupling free energy is given by $\Delta \Delta G_{\text{int}} = -RT \ln[(\theta_{\text{ww}} \theta_{\text{mm}})/(\theta_{\text{wm}} \theta_{\text{mw}})]$, where θ is the gating equilibrium constant (Table 1) for wild-type (ww), single mutant (wm, mw) or double mutant (mm) receptors.

charged residues at the end of β -strand 10 of the C-loop are generally important for receptor activation.

The proposed coupling pathway also contains Val 46, Pro 272 and Ser 269 (Fig. 1). To look for energetic coupling, we mutated each residue individually and in combination and measured channel gating equilibrium constants as above. Because coupling may extend beyond residue pairs to residue triads, however, we cast the experimental results as three-dimensional mutant cycles (Fig. 3). Reducing the side chain volume of either Val 46 or Pro 272 markedly impairs channel gating: V46A reduces the gating equilibrium constant 470-fold, and P272G reduces it 150-fold (Fig. 3a, top plane, and Table 1). When the two mutations are combined, however, gating is about 20-fold more efficient than expected from the sum of the two mutations. Analysis of the mutant cycle reveals a substantial coupling energy between Val 46 and Pro 272 of $-1.6 \pm 0.07 \text{ kcal mol}^{-1}$.

Coupling between Val 46 and Pro 272 remains strong in receptors in which Glu 45 is mutated to alanine (Fig. 3a, bottom plane): the coupling energy increases further to $-2.4 \pm 0.07 \text{ kcal mol}^{-1}$, presumably because release of Glu 45 from Arg 209 enables Val 46 to interact more strongly with Pro 272. Notably, when the smaller alanine residue is substituted for both Glu 45 and Val 46, binding is almost uncoupled from gating, and mutation of Pro 272 has no further effect (Fig. 3a, lower right limb, and Table 1). Thus, Val 46 and Glu 45 are the primary partners for Pro 272, indicating an energetically coupled triad. We sought further evidence for a coupled triad by examining the cycle comprising mutations of Glu 45 and Pro 272 (Fig. 3a, left plane). The coupling energy for this pair is modest, $-0.7 \pm 0.07 \text{ kcal mol}^{-1}$, but it increases to $-1.4 \pm 0.07 \text{ kcal mol}^{-1}$ in receptors in which Val 46 is mutated to alanine (Fig. 3a, right plane). Thus, Val 46, Glu 45 and Pro 272 form an energetically coupled triad, probably through hydrophobic contacts.

To look for energetic coupling near the channel-forming M2 helix^{1,18}, we mutated residues that emerge from M2 to form the cavity accommodating Val 46 (Fig. 1). Mutations of Ala 270, Val 271 and Leu 273 do not alter channel gating (Supplementary Table S1), but mutation of Ser 269 to leucine enhances gating around threefold (Table 1). Thus, we tested for energetic coupling between Ser 269 and the spatially proximal Val 46 (Fig. 3b, top plane). When mutations of Val 46 and Ser 269 are combined in a single receptor, gating is about 20-fold more efficient than expected from the sum of the individual mutations. Mutant cycle analysis yields a substantial coupling energy between Val 46 and Ser 269 of $-1.7 \pm 0.07 \text{ kcal mol}^{-1}$ (Fig. 3b, top plane), probably arising through hydrophobic interactions.

We next tested whether Val 46, Ser 269 and Pro 272 form an energetically coupled triad, again by studying pairs of mutations in the presence of a third mutation. In receptors in which Pro 272 is mutated to glycine, the coupling energy between Val 46 and Ser 269 decreases to $0.4 \pm 0.05 \text{ kcal mol}^{-1}$ (Fig. 3b, bottom plane), suggesting a coupled triad. The cycle comprising mutations of Ser 269 and Pro 272 shows a negligible coupling energy of $-0.1 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3b, left plane), but when the intervening Val 46 is mutated to alanine, the coupling energy increases to $1.9 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3b, right plane). Thus, Val 46, Ser 269 and Pro 272 form an energetically coupled triad.

Finally, we examined energetic coupling between spatially separate residues with the expectation that these residues would not show coupling. The cycle comprising mutations of Glu 45 and Val 46 shows a negligible coupling energy of $-0.1 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3c, front plane, and Table 1). Thus, although Glu 45 and Val 46 are consecutive in the primary sequence, they are independent because they contribute to distinct limbs of the principal pathway. Glu 45 and Val 46 remain essentially independent in receptors in which Ser 269 is mutated to leucine, showing a low coupling energy of $-0.25 \pm 0.07 \text{ kcal mol}^{-1}$ (Fig. 3c, back plane, and Table 1).

The cycle comprising mutations of Glu 45 and Ser 269 shows a weak coupling energy of $-0.4 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3c, left plane), approaching independence, and these residues become independent

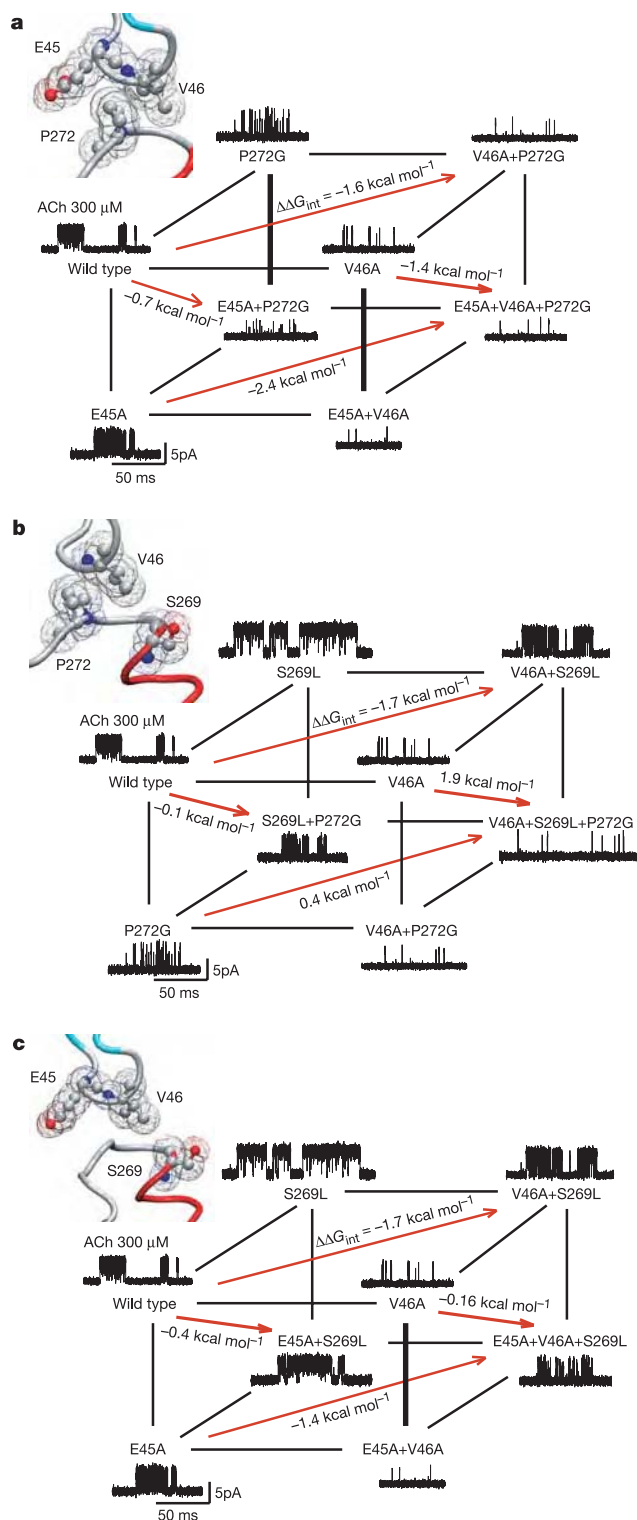


Figure 3 | Energetic coupling between residues in the α -subunit. Structures are derived from PDB 2BG9. Three-dimensional mutant cycles are shown with single-channel currents elicited by $300 \mu\text{M}$ ACh ($30 \mu\text{M}$ for αS269L) next to each receptor type. Coupling free energies for several cycle planes are highlighted (orange arrows). Channel gating rate and equilibrium constants for each receptor type are given in Table 1, and the complete sets of fitted binding and gating rate constants are given in Supplementary Table S2. **a**, Energetic coupling among Glu 45, Val 46 and Pro 272. Coupling free energies for the front and back planes are -0.1 and $0.8 \text{ kcal mol}^{-1}$, respectively (Table 1). **b**, Energetic coupling among Val 46, Pro 272 and Ser 269. Coupling free energies for the front and back planes are -1.6 and $0.4 \text{ kcal mol}^{-1}$, respectively (Table 1). **c**, Energetic coupling among Glu 45, Val 46 and Ser 269. Coupling free energies for the front and back planes are -0.1 and $-0.25 \text{ kcal mol}^{-1}$, respectively (Table 1).

in receptors in which Val 46 is mutated to alanine, showing a coupling energy of $-0.16 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3c, right plane). Thus, Glu 45 and Ser 269 are independent because they contribute to distinct limbs of the principal pathway. Val 46 and Ser 269, which show a coupling energy of $-1.7 \text{ kcal mol}^{-1}$ (Fig. 3c, top plane), remain coupled in receptors in which Glu 45 is mutated to alanine, showing a coupling energy of $-1.4 \pm 0.06 \text{ kcal mol}^{-1}$ (Fig. 3c, bottom plane). Thus, coupling between Val 46 and Ser 269 is independent of Glu 45 because Glu 45 extends away from this pair.

A key issue is whether the coupling energies measured reflect true interaction energies or conformational rearrangements secondary to mutation. Each single mutation probably causes a rearrangement of neighbouring residues, but the energy associated with each rearrangement cancels in the overall mutant cycle¹⁹. Crystal structures of single and double mutations of barnase showed local rearrangements after mutation of each member of an interacting pair, but these rearrangements were also present in the structure of the double mutation, and mutant cycle analysis identified a true interaction energy¹³. Complications arise if the local perturbation spreads between mutational sites to cause rearrangements in the double mutation that are not seen in the single mutations. If such secondary rearrangements were responsible for the coupling energies observed here, however, mutant cycle analysis of non-contact residues would be expected to show apparent coupling. Our collective findings show that when two or three residues make close contact in the structural model¹ they are energetically coupled, but when they are spatially separate they are independent. Thus, the coincidence between physical proximity and energetic coupling constitutes powerful evidence that the observed coupling energies reflect genuine interactions and that the principal pathway in Fig. 1 links binding to gating.

The above studies focus on the α -subunit, which forms the principal face of the agonist-binding site²⁰. Several residues of the proposed coupling pathway are also found in β -, δ - and ϵ -subunits, but these subunits do not form the principal face of a binding site. To assess subunit specificity of the key coupling residues, we tested whether equivalent residues in the non α -subunits contribute to channel gating. Mutation of any of the three non α -subunits at positions equivalent to Glu 45, Val 46, Pro 272 or Ser 269 leads to gating in the normal range, unlike equivalent mutations in the α -subunit (Supplementary Table S2). Similarly, the α -subunits change conformation on brief exposure to ACh, whereas the non

α -subunits do not⁶. Thus, agonist binding couples to channel gating through the principal pathway in the α -subunit.

Our results identify the primary pathway that couples agonist binding to channel gating in the muscle nicotinic receptor and provide a template for understanding coupling in other members of the Cys-loop receptor superfamily. The structural model of the *Torpedo* receptor¹ provided the first step in identifying the link between agonist binding and channel gating. We have now established functional coupling among residues in the static structural model by applying knowledge of residue conservation, structures known to change conformation in response to agonist, direct measures of channel gating and residue-by-residue analysis of energetic coupling.

The primary coupling pathway draws together distinct regions from the binding and pore domains to form an interdependent network. The concept of an interdependent network had been established^{2,3}, but the principal pathway remained elusive because of the lack of a structural model to trace the pathway. Two other loops from the binding domain, the Cys-loop and $\beta 8$ – $\beta 9$ linker, are also essential for coupling binding to gating^{21–25}, and may act as inputs to the primary pathway. Both loops lodge in the hydrophobic core of the subunit, with His 134 of the Cys-loop and Glu 175 and Trp 176 of the $\beta 8$ – $\beta 9$ linker flanking the Arg 209/Glu 45 salt bridge on opposite sides (see Protein Data Bank accession code 2BG9). However, these local inputs make minor or modest contributions to the principal pathway: as compared with R209Q, which impairs channel gating 46-fold, W176F has no effect, E175Q impairs it 3-fold and H134A impairs it 4-fold (Supplementary Table S1). The Cys-loop and $\beta 8$ – $\beta 9$ linker make additional distributed contacts in which the Cys-loop contacts β -strand 10 and the M2–M3 linker, and the $\beta 8$ – $\beta 9$ linker contacts β -strands 1 and 10, but energetic coupling between these regions has not been established. Disease-causing mutations have been identified in residues of the primary pathway^{14,26}, the Cys-loop²⁷ and $\beta 8$ – $\beta 9$ linker²⁸, underscoring their functional significance.

In summary, capping of the agonist-binding site by the C-loop^{4,7} is conveyed to the M2 pore helix through a salt bridge in series with a precise assembly of hydrophobic side chains. The salt bridge is found in all receptors of the Cys-loop superfamily and is located in the hydrophobic core of the subunit where its electrostatic force is maximized. Residues linked to the salt bridge form pairs and triads that are stabilized by close fit and hydrophobicity to join inner and peripheral β -sheets; these residues are not conserved across the

Table 1 | Channel gating rate and equilibrium constants and coupling free energies

α -Subunit mutations	β (s^{-1})	α (s^{-1})	θ (β/α)	$\Delta\Delta G_{\text{int}}$ (kcal mol $^{-1}$)
Wild type	44,000	1,700	27	–
R209Q	9,400	16,000	0.59	–
E45K	5,110	22,800	0.224	–
E45R	21,000	5,100	4.1	–
E45A	29,000	14,000	2.1	–
V46A	155	2,700	0.057	–
P272G	1,300	7,400	0.17	–
S269L	81,200	980	83	–
E45R + R209E	56,000	2,900	20	–
E45R + R209Q	41,000	2,200	19	-3.1 ± 0.10
E45A + V46A	78	17,000	0.0046	-0.1 ± 0.06
E45A + S269L	75,000	5,400	14	-0.4 ± 0.06
E45A + P272G	820	20,000	0.042	-0.7 ± 0.07
V46A + S269L	5,900	2,000	2.9	-1.7 ± 0.07
V46A + P272G	38	6,800	0.0056	-1.6 ± 0.07
S269L + P272G	6,000	9,100	0.66	-0.1 ± 0.06
E45A + V46A + S269L	3,100	9,600	0.32	$-1.4 \pm 0.06^*$, $-0.16 \pm 0.06^\dagger$, $-0.25 \pm 0.07^\ddagger$
E45A + V46A + P272G	61	11,000	0.0054	$-2.4 \pm 0.07^\S$, $-1.4 \pm 0.07^\parallel$, $0.8 \pm 0.08^\P$
V46A + S269L + P272G	110	9,200	0.011	$1.9 \pm 0.06^\#$, $0.4 \pm 0.05^\star$, $0.4 \pm 0.06^{**}$

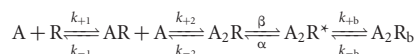
Channel opening (β) and closing (α) rate constants were estimated by global fitting of open and closed dwell times, obtained over a range of ACh concentrations, to a mechanistic description of receptor activation (Methods). Standard errors of the gating rate and equilibrium constants, the ACh association and dissociation rate constants, and the range of ACh concentrations are given in Supplementary Table S1. Standard errors are given for the coupling free energies, which are considered significantly different from zero when they exceed twice the standard error, assuming a normal distribution and a two-sided P value of 0.05. The mutant cycle starts from *E45A (Fig. 3c), † V46A (Fig. 3c), ‡ S269L (Fig. 3c), § E45A (Fig. 3a), $^\parallel$ V46A (Fig. 3a), ¶ P272G (Fig. 3a), $^\#$ V46A (Fig. 3b), * P272G (Fig. 3b), ** S269L (Fig. 3b).

Cys-loop receptor superfamily, but are conserved in ACh, GABA_A, glycine and 5-HT₃ receptor subfamilies. Thus, the primary coupling pathway integrates contributions from several structural domains, with its distal limb likely representing the point at which the binding domain triggers opening of the channel.

METHODS

Construction of mutant ACh receptor subunits, their expression in BOSC 23 cells, measurements of ¹²⁵I-labelled α-bungarotoxin binding, and recording of single-channel currents are described in the Supplementary Methods.

Single channel kinetic analysis. The kinetics of activation for wild-type and mutant receptors were analysed as described¹¹. In brief, open and closed dwell times, obtained over a range of ACh concentrations, were simultaneously fitted with MIL software (QUB suite, State University of New York, Buffalo) using the following mechanistic description:



where A is ACh, R is the receptor, k_{+n} and k_{-n} are agonist association and dissociation rate constants, β and α are channel opening and closing rate constants, and k_{+b} and k_{-b} are channel blocking and unblocking rate constants. Data were obtained at half-log unit intervals over a 30- to 300-fold range of ACh concentrations, with dwell times from two to three patches per ACh concentration fitted simultaneously. The number of open and closed intervals per patch ranged from 4,000 to 8,000. Error estimates of the fitted parameters were determined by MIL as described³⁰. The resulting channel opening and closing rate constants were used to obtain the channel gating equilibrium constants (Table 1).

Received 6 May; accepted 9 August 2005.

- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Kash, T., Jenkins, A., Kelly, J., Trudell, J. & Harrison, N. L. Coupling of agonist binding to channel gating in the GABA_A receptor. *Nature* **421**, 272–275 (2003).
- Bouzat, C. *et al.* Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel. *Nature* **430**, 896–900 (2004).
- Gao, F. *et al.* Agonist-mediated conformational changes in acetylcholine-binding protein revealed by simulation and intrinsic tryptophan fluorescence. *J. Biol. Chem.* **280**, 8443–8451 (2005).
- Law, R., Henchman, R. & McCammon, J. A. A gating mechanism proposed from a simulation of a human α7 nicotinic acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **102**, 6813–6818 (2005).
- Unwin, N., Miyazawa, A., Li, J. & Fujiyoshi, Y. Activation of nicotinic acetylcholine receptor involves a switch in conformation of the α subunits. *J. Mol. Biol.* **319**, 1165–1176 (2002).
- Celie, P. H. N. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
- Colquhoun, D. & Sakmann, B. Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels. *Nature* **294**, 464–466 (1981).
- Sine, S. M., Claudio, T. & Sigworth, F. J. Activation of *Torpedo* acetylcholine receptors expressed in mouse fibroblasts. Single channel current kinetics reveal distinct agonist binding affinities. *J. Gen. Physiol.* **96**, 395–437 (1990).
- Grosman, C., Zhou, M. & Auerbach, A. Mapping the conformational wave of acetylcholine receptor channel gating. *Nature* **403**, 773–776 (2000).
- Lee, W. Y. & Sine, S. M. Invariant aspartic acid in muscle nicotinic receptor contributes selectively to the kinetics of agonist binding. *J. Gen. Physiol.* **124**, 555–567 (2004).
- Horovitz, A. & Fersht, A. Strategy for analyzing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.* **214**, 613–617 (1990).
- Vaughan, C., Harryson, P., Buckle, A. & Fersht, A. A structural double-mutant cycle: estimating the strength of a buried salt bridge in barnase. *Acta Crystallogr. D* **58**, 591–600 (2002).
- Castaldo, P. *et al.* A novel hyperekplexia-causing mutation in the pre-transmembrane segment 1 of the human glycine receptor α1 subunit reduces membrane expression and impairs gating by agonists. *J. Biol. Chem.* **279**, 25598–25604 (2004).
- Vicente-Agullo, F. *et al.* Multiple roles of the conserved key residue arginine 209 in neuronal nicotinic receptors. *Biochemistry* **40**, 8300–8306 (2001).
- Kash, T. L., Dizon, M. J., Trudell, J. R. & Harrison, N. L. Charged residues in the β2 subunit involved in GABA_A receptor activation. *J. Biol. Chem.* **279**, 4887–4893 (2004).
- Hu, X. Q., Zhang, L., Stewart, R. R. & Weight, F. F. Arginine 222 in the pre-transmembrane domain 1 of 5-HT_{3A} receptors links agonist binding to channel gating. *J. Biol. Chem.* **278**, 46583–46589 (2003).
- Wilson, G. G. & Karlin, A. The location of the gate in the acetylcholine receptor channel. *Neuron* **20**, 1269–12681 (1998).
- Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **29**, 9343–9352 (1990).
- Sine, S. M. The nicotinic receptor ligand binding domain. *J. Neurobiol.* **3**, 431–446 (2002).
- Lyford, L. K., Sproul, A. D., Eddins, D., McLaughlin, J. T. & Rosenberg, R. L. Agonist-induced conformational changes in the extracellular domain of α7 nicotinic acetylcholine receptors. *Mol. Pharmacol.* **64**, 650–658 (2003).
- Schofield, C. M., Jenkins, A. & Harrison, N. L. A highly conserved aspartic acid residue in the signature disulfide loop of the α1 subunit is a determinant of gating in the glycine receptor. *J. Biol. Chem.* **278**, 34079–34083 (2003).
- Absalom, N. L., Lewis, T. M., Kaplan, W., Pierce, K. D. & Schofield, P. R. Role of charged residues in coupling ligand binding and channel activation in the extracellular domain of the glycine receptor. *J. Biol. Chem.* **278**, 50151–50157 (2003).
- Chakrapani, S., Bailey, T. D. & Auerbach, A. Gating dynamics of the acetylcholine receptor extracellular domain. *J. Gen. Physiol.* **123**, 341–356 (2004).
- Sala, F., Mulet, J., Sala, S., Gerber, S. & Criado, M. Charged amino acids of the N-terminal domain are involved in coupling binding and gating in α7 nicotinic receptors. *J. Biol. Chem.* **280**, 6642–6647 (2005).
- Croxen, R. *et al.* Mutations in different functional domains of the human muscle acetylcholine receptor α subunit in patients with the slow-channel congenital myasthenic syndrome. *Hum. Mol. Gen.* **6**, 767–774 (1997).
- Shen, X. M. *et al.* Mutation causing severe myasthenia reveals functional asymmetry of AChR signature cysteine loops in agonist binding and gating. *J. Clin. Invest.* **111**, 497–505 (2003).
- Sine, S. M. *et al.* Naturally-occurring mutations at the acetylcholine receptor binding site independently alter ACh binding and channel gating. *J. Gen. Physiol.* **120**, 483–496 (2002).
- Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).
- Qin, F., Auerbach, A. & Sachs, F. Estimating single channel kinetic parameters from idealized patch clamp data containing missed events. *Biophys. J.* **70**, 264–280 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. R. Free for contributions to the molecular biology and T. Therneau for help with statistical analysis. This work was supported by a grant from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M.S. (sine@mayo.edu).

LETTERS

Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel

Sarah C. R. Lummis^{1*}, Darren L. Beene^{2*}, Lori W. Lee², Henry A. Lester³, R. William Broadhurst¹
& Dennis A. Dougherty²

5-Hydroxytryptamine type 3 (5-HT₃) receptors are members of the Cys-loop receptor superfamily¹. Neurotransmitter binding in these proteins triggers the opening (gating) of an ion channel by means of an as-yet-uncharacterized conformational change. Here we show that a specific proline (Pro 8*), located at the apex of the loop between the second and third transmembrane helices (M2–M3)^{2,3}, can link binding to gating through a *cis*–*trans* isomerization of the protein backbone. Using unnatural amino acid mutagenesis, a series of proline analogues with varying preference for the *cis* conformer was incorporated at the 8* position. Proline analogues that strongly favour the *trans* conformer produced non-functional channels. Among the functional mutants there was a strong correlation between the intrinsic *cis*–*trans* energy gap of the proline analogue and the activation of the channel, suggesting that *cis*–*trans* isomerization of this single proline provides the switch that interconverts the open and closed states of the channel. Consistent with this proposal, nuclear magnetic resonance studies on an M2–M3 loop peptide reveal two distinct, structured forms. Our results thus confirm the structure of the M2–M3 loop and the critical role of Pro 8* in the 5-HT₃ receptor. In addition, they suggest that a molecular rearrangement at Pro 8* is the structural mechanism that opens the receptor pore.

The neurotransmitter binding site in the Cys-loop superfamily of ion channels is located about 60 Å from the channel pore, presenting a conundrum as to the molecular events that link binding and gating^{1,3}. Evidence implicates the M2–M3 loop, and in low-resolution structural studies this region interacts with loops 2 and 7 of the extracellular domain (Fig. 1)^{4–8}. In the cation-selective nicotinic acetylcholine (nACh) and 5-HT₃ receptors, the apex of the M2–M3 loop contains a conserved proline (Pro 8*) that is ideally placed to provide a hinge for movement of the channel-lining M2 helix. This proline is essential for receptor function in the 5-HT₃ receptor. We found that replacement of Pro 8* with Gly, Ala, Cys, Val, Lys or Asn gave receptors that were trafficked to the membrane properly and displayed wild-type binding properties for a radio-labelled antagonist (Supplementary Information); however, all receptors were non-functional². This indicates that mutations at Pro 8* affect receptor function by altering gating, not ligand binding or folding/assembly, and that a unique characteristic of proline is absolutely required for the gating of the receptor.

To explore the possible role of Pro 8* in channel gating, we used nonsense suppression in *Xenopus* oocytes to incorporate a series of proline analogues (Fig. 2a) at position 8* of the 5-HT₃ receptor⁹. This technique involves inserting a nonsense codon (TAG) in place of the proline codon into the DNA encoding the receptor subunit, synthesizing messenger RNA, and injecting this into *Xenopus* oocytes along with a transfer RNA that both recognizes the TAG codon and

has the appropriate proline analogue attached at the 3' end. The resulting mutant receptors were examined using voltage-clamp analysis of 5-HT-induced whole-cell responses and immunofluorescence.

When incorporated into a protein, proline disrupts main-chain hydrogen bonding, because it lacks a backbone NH moiety¹⁰.

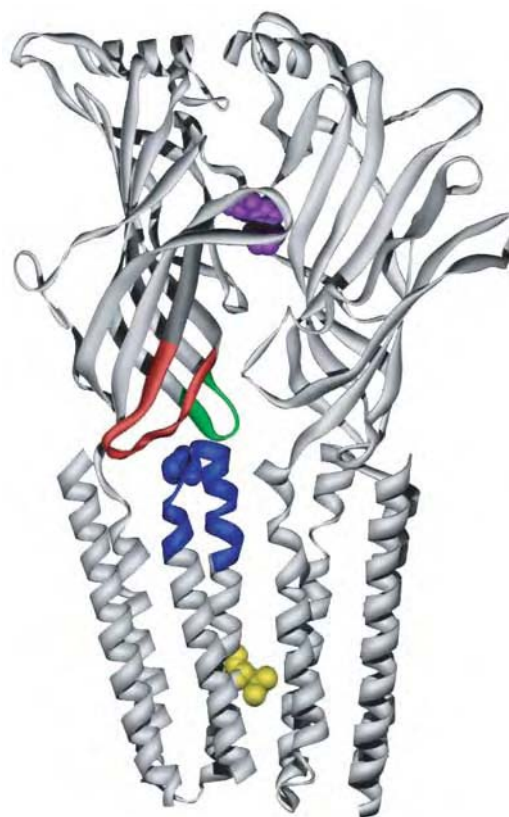


Figure 1 | Overall layout of the 5-HT₃ receptor showing the extracellular (predominantly β -sheet) and transmembrane (α -helical) regions. Shown are two adjacent subunits from a homology model of the 5-HT₃ receptor³⁰ built from AChBP (extracellular domain)¹⁹ and the nACh receptor transmembrane domain (transmembrane domain)¹⁵. The binding site Trp is purple; loop 7 is red; loop 2 is green; the conserved Leu thought to contribute to the gate is yellow (space filling); and strand β 7 is dark grey. The region corresponding to the peptide studied by NMR (Fig. 4), comprising the M2–M3 loop and parts of M2 and M3, is shown in blue, with Pro 8* in space-filling model.

¹Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK. ²Division of Chemistry and Chemical Engineering, and ³Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.

*These authors contributed equally to this work.

Previously, we probed a highly conserved Pro in the middle of the first transmembrane helix (M1) by substitution with an α -hydroxy acid such as Vah (α -hydroxyvaline; Fig. 2a), which similarly lacks an NH^+ . In M1 Vah substitutes effectively for Pro; however, at position 8* Vah produces non-functional receptors. Similarly, the *N*-methyl amino acids Sar (sarcosine) and *N*-Me-Leu give non-functional receptors.

Proline is the only natural cyclic amino acid. Incorporation of nine cyclic proline analogues at position 8* yielded mixed results: some produced functional receptors and others did not (Fig. 2a). Thus, neither the cyclic structure nor the lack of a hydrogen bond donor are key to the proper functioning of Pro at 8*.

Proline is also unusual in forming *cis* peptide bonds at a frequency ($\sim 5\%$) much higher than any other naturally occurring amino acid ($<0.1\%$)¹², reflecting an intrinsic diminution of the energy gap between *cis* and *trans* forms in proline¹⁰. The analogues 2,4-CH₂-Pro and 2-Me-Pro show a reduced preference for a *cis* conformation relative to Pro¹³, and these produced non-functional receptors (that

is, no current was seen in response to normally saturating concentrations of serotonin). Immunocytochemical experiments revealed that these receptors did reach the cell surface (Supplementary Information).

Of the proline analogues that give functional receptors, *t*-4F-Pro and *c*-4F-Pro give near-wild-type EC_{50} (effector concentration for half-maximum response) values (1.38 ± 0.06 and $1.15 \pm 0.12 \mu\text{M}$, respectively), and also show intrinsic *cis-trans* preferences similar to native proline. Other analogues, however, show much greater propensities to adopt the *cis* conformer. Most extreme is Dmp (5,5-dimethylproline), which strongly favours the *cis* form (Table 1). Incorporation of this amino acid at position 8* resulted in a 60-fold decrease in EC_{50} . Furthermore, in oocytes that expressed Dmp-containing receptors, agonist exposure led to partially irreversible inactivation, as evidenced by a continual increase in standing current after initial receptor activation; this current could be substantially reversed by the addition of the channel blocker TMB-8 (ref. 11) (Fig. 3). Dmp and other analogues with high *cis* preference showed low and variable expression levels in *Xenopus* oocytes. This could reflect difficulties in folding the protein when a key proline is present in an unconventional conformation. Overall, we found three types of behaviours for incorporation of proline analogues. Pro derivatives strongly favouring *trans* produced no responses to agonist but did express and were inserted into the membrane. Intermediate derivatives produced responses with a range of EC_{50} values. The most *cis*-producing residue produced highly sensitive receptors that displayed partially irreversible activation.

An evaluation of a series of Pro analogues that exhibit a range of intrinsic *cis-trans* energy gaps revealed a remarkable trend. As shown in Fig. 2b, a linear relationship was observed between the *cis-trans* energy gap of the proline analogue ($\Delta\Delta G(c-t)$) and activation of the receptor.

EC_{50} , the functional output we measure, reflects a composite of binding and gating events. However, as noted above, substitution at Pro 8* by natural amino acids does not change antagonist binding affinity. In addition, the apparent affinity of the 5-HT₃ receptor competitive antagonist MDL72222 is not significantly altered when Pro 8* is mutated to Pip (pipecolic acid) or Dmp (see Supplementary Information). These data show that for amino acids with a wide range of *cis* preferences (<0.1 –71%) at Pro8*, the binding site is unperturbed. We therefore consider the variations in EC_{50} to reflect changes in receptor gating, not binding¹⁴, justifying the free energy treatment of Fig. 2b.

The slope of the line in Fig. 2b is 1, within experimental error. This indicates that the full energetic perturbation of the proline *cis-trans* equilibrium induced by the mutation is felt in the gating equilibrium. Together, these observations provide a compelling link between the conformational isomerization at residue 8* and the gating of the 5-HT₃ receptor.

To provide further support that a proline in the M2-M3 loop can serve as a structural switch, we used nuclear magnetic resonance

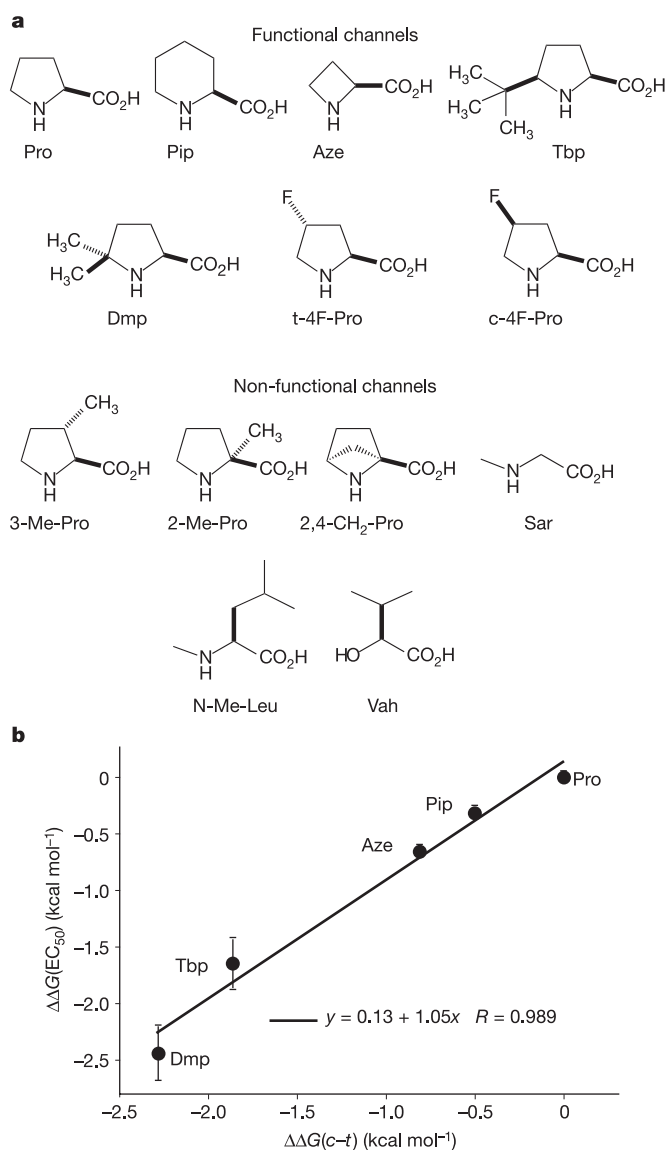


Figure 2 | Unnatural amino acid mutagenesis. **a**, Residues used in this study. As noted in the text, the conventional amino acids Gly, Ala, Cys, Val, Lys and Asn also give non-functional channels. **b**, Linear energy correlation between *cis-trans* isomerization and receptor activation. Data are from Table 1. Error bars (s.e.m.) are plotted for all data points, but for Pro, Pip and Aze they are partially obscured by the marker.

Table 1 | Influence of proline isomerism on receptor function

Residue	Per cent <i>cis</i> *	EC_{50} (μM)†	$\Delta\Delta G(c-t)$ (kcal mol^{-1})‡	$\Delta\Delta G(\text{EC}_{50})$ (kcal mol^{-1})§
Pro	5	1.29 ± 0.07	0	0
Pip	12	0.75 ± 0.06	-0.54	-0.32
Aze	18	0.42 ± 0.03	-0.85	-0.66
Tbp	55	0.030 ± 0.024	-1.86	-1.73
Dmp	71	0.021 ± 0.009	-2.28	-2.47

* Determined from studies of model peptide systems reported previously^{21–23}. Because variations are seen depending on methodology and exact model system, all values are referenced to the *cis-trans* ratio seen for Pro in the same study, and Pro is set to 5%, the value obtained from statistical surveys of protein structures.

† Values are $\pm$ s.e.m.

‡ Values are relative to proline.

§ Equals $-\text{RT}(\ln(\text{EC}_{50}(\text{mutant})/\text{EC}_{50}(\text{Pro})))$.

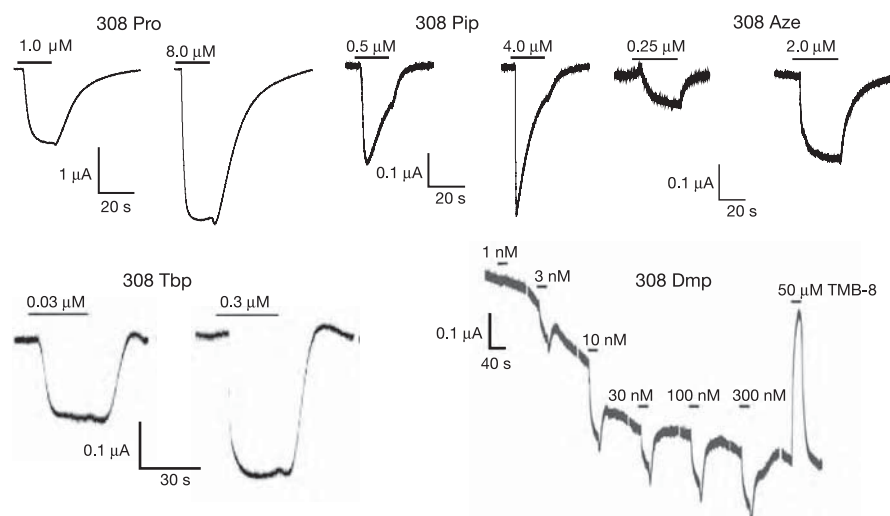


Figure 3 | Current traces in voltage-clamp experiments (at approximately EC_{50} and approximately I_{max}) for receptors expressing unnatural amino acids. For Dmp at position 8* (residue 308 of 5-HT₃), a full dose-response

range is given along with the TMB-8 block of standing currents seen after receptor activation. Bars represent 5-HT application.

(NMR) spectroscopy to evaluate the structure of a 20-amino-acid peptide spanning the M2–M3 loop (SDTLPATAIGTPLIGVYFVV; Pro 8* is italic). When solubilized in a micellar solution of SDS, this relatively short sequence is structured, with two segments flanking Pro 8* displaying a significant amount of α -helical character (residues 6–10 and 13–20 of the peptide; see Supplementary Information). Thus, key features of the M2–M3 loop, as reported in the low-resolution nACh receptor structure^{8,15} (see also Fig. 1), are recapitulated in the model peptide. Interestingly, two conformers are clearly evident in the peptide structure, with the major form about five times more prevalent than the minor form (Fig. 4). In the major conformer Pro 8* was assigned the *trans* conformation based on observed nuclear Overhauser effect connections. Some features of the minor form cannot be resolved due to overlapping resonances,

but the TOCSY (total correlated spectroscopy) spectrum clearly shows that the two glycines that flank Pro 8* (positions 10 and 15 in the peptide; 6* and 11* in the protein) are in different environments in the two conformers, with both pairs of diastereotopic H $^{\alpha}$ nuclei giving separate resonances in the minor conformer (Fig. 4). These results are consistent with two conformations at Pro 8* producing two conformers of the peptide.

We propose the following mechanism for gating in the 5-HT₃ receptor. In the closed state of the channel, Pro 8* is in the *trans*

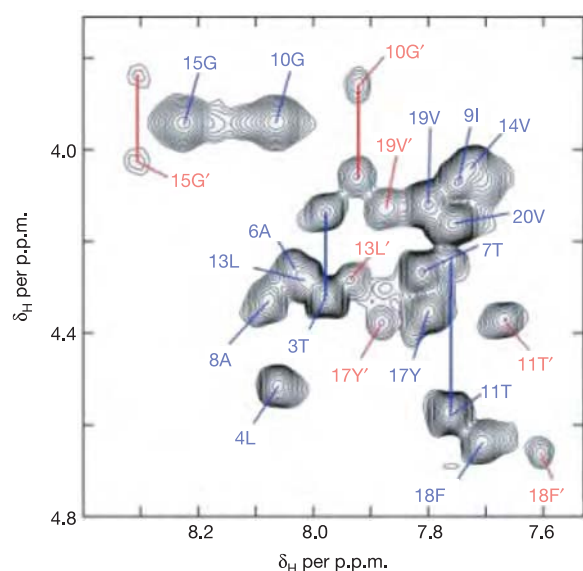


Figure 4 | ^1H NMR of a model peptide. The H^N–H $^{\alpha}$ fingerprint region of a TOCSY spectrum for the 20-residue peptide discussed in the text resolves signals for the major (blue) and minor (red) conformers in a micellar solution of SDS at 298 K. Integration indicates that 13–20% of the minor conformer is present. Signals arising from distinct H $^{\alpha}$ or H $^{\beta}$ environments within a particular residue are connected with vertical lines.

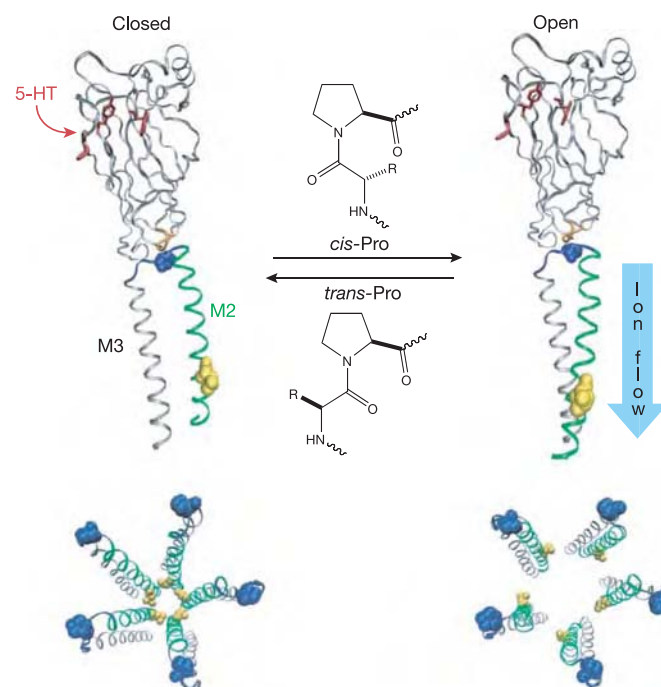


Figure 5 | A proposed gating mechanism. A single 5-HT₃ receptor subunit is depicted, illustrating how *trans*–*cis* isomerization at Pro 8* (blue space-filling model) could function as a hinge for movement of M2 (green) during gating, moving the occluding Leu 9' residue (yellow) away from the channel region. The open structure on the right was generated by manually converting Pro 308 to the *cis* conformation and treating M2 as a rigid body. Top: side view; bottom: top view, showing only M2, M3 and the M2–M3 loop.

conformation. Binding of ligand initiates a structural change that culminates in isomerization of Pro 8* to the *cis* form. This change at a crucial pivot point reorients the M2 transmembrane helix, opening the channel. We have crudely modelled the change that would occur if Pro 8* isomerized from the *trans* to the *cis* conformation (Fig. 5), treating the M2 α -helix as a rigid body. Because of the hinge location of Pro 8*, the structurally modest *cis-trans* isomerization at this single residue propagates to a substantial structural effect more than adequate to gate the channel.

Our model also suggests a natural coupling between the agonist binding site and the conformational change associated with gating. Several workers have noted the proximity of loops 2 and 7 (the eponymous Cys loop) to the M2–M3 loop^{4–8,16}. Although the image of Fig. 1 is not based on atomic-resolution data, Pro 8* in its position at the apex of M2–M3 is in close proximity to both loops 2 and 7. It has been proposed that loops 2 and 7 'lock' the M2 helix in the closed conformation⁸. Expanding upon this model, we propose that loops 2 and 7 act as a caliper to immobilize Pro 8* in a *trans* conformation. Arrival of agonist causes a movement of the key binding site tryptophan (Trp 183 in the 5-HT₃ receptor; Trp 149 in the nACh receptor)^{17–20}. This Trp is directly linked to loop 7 by a six-amino-acid stretch of sheet structure termed $\beta 7$ (dark grey in Fig. 1). Thus, the binding site Trp may be the actuator that, perhaps via $\beta 7$, releases the clamp on M2–M3. Pro 8* can then undergo a spontaneous (or protein aided) *cis-trans* isomerization, gating the channel.

The per cent *cis* values of Table 1 are based on model peptide systems^{21–23}; in any particular context the equilibrium could be perturbed. It may well be that structural features of the wild-type receptor act to favour the *cis* conformer of the native proline at position 8*, once the structural change associated with agonist binding has occurred. There is also an issue of timescale. Estimates of the opening rate for the 5-HT₃ receptor are in the 10–100 s^{–1} range^{24,25}. Intrinsic prolyl *cis-trans* isomerization rates in peptides are within a factor of ten of this regime, and structural features of the receptor could act to accelerate significantly the isomerization rate. Protein prolyl isomerases are well known, and even a simple hydrogen bond to the proline amide nitrogen can accelerate isomerization ~260-fold²⁶. It is possible that the receptor has incorporated structural features that facilitate *cis-trans* isomerization. Therefore, prolyl *cis-trans* isomerization is a kinetically viable candidate for the gating switch.

We have established a clear correlation between the conformational preference of a single, specific proline residue and the gating efficiency of a Cys-loop receptor. Proline has unique structural and conformational properties, and it is found with anomalously high frequency in the transmembrane regions of ion channels and transporters, suggesting a key role in structural changes associated with transmembrane signalling^{27,28}. The unnatural amino acid mutagenesis approach has allowed an exquisitely detailed probe of the Pro 8* site, yielding an experimentally based model of channel opening involving a precise structural change at the amino acid level that induces gating. The model is consistent with all available structural data, and it suggests much future work. It will be interesting to perform the same 'proline scan' on the nACh receptor, which also contains a proline at position 8*, but which displays a much faster opening rate. In contrast, other Cys-loop receptors do not contain this proline, and the possibility of a *cis-trans* isomerization at a non-proline site or a completely different gating mechanism should be explored.

METHODS

Mutagenesis and preparation of cRNA and oocytes. Mutant 5-HT_{3A} receptor subunits were developed using pcDNA3.1 (Invitrogen) containing the complete coding sequence for the 5-HT_{3A(b)} subunit from mouse neuroblastoma N1E-115 cells as previously described²⁹. For nonsense suppression, the proline codon at 308 was replaced by TAG as previously described⁹. Wild type and mutant receptor subunit coding sequences were then subcloned into pGEMHE. This

was linearized with *Nhe*I (New England Biolabs) and cRNA synthesized using the T7 mMESSAGE mMACHINE kit (Ambion). Oocytes from *Xenopus laevis* were prepared and maintained as described previously⁹.

Synthesis of tRNA and dCA-amino acids. Unnatural amino acids were chemically synthesized as nitroveratryloxycarbonyl (NVOC) protected cyanomethyl esters and coupled to the dinucleotide dCA, which was then enzymatically ligated to 74-mer THG73 tRNA_{CUA} as detailed previously⁹. Immediately before co-injection with mRNA, tRNA-aa was deprotected by photolysis. Typically, 5 ng mRNA and 25 ng tRNA-aa were injected into stage V–VI oocytes in a total volume of 50 nl. For control experiments, mRNA was injected in the absence of tRNA, and with the THG73 74-mer tRNA. Experiments were performed 18–36 h after injection.

Characterization of mutant receptors. 5-HT-induced currents were recorded from individual oocytes using two-electrode voltage clamp with either a GeneClamp 500 amplifier or an OpusXpress system (Molecular Devices Axon Instruments). All experiments were performed at 22–25 °C. Serotonin (creatinine sulphate complex, Sigma) was stored as 25 mM aliquots at –80 °C, diluted in calcium-free ND96, and delivered to cells via computer-controlled perfusion systems. The holding potential was –60 mV unless otherwise specified. EC₅₀ data were obtained from at least six independent experiments using oocytes from at least three different batches. Further details are available in Supplementary Information.

NMR. The NMR sample was prepared containing 1 mM peptide, 200 mM sodium dodecyl-d25 sulphate, 150 mM sodium chloride, 20 mM sodium phosphate, 1 mM EDTA, 20 μ M 3,3,3-trimethylsilylpropionate and 10% D₂O at pH 6.0 to a final volume of 550 μ l in a 5-mm Ultra-Imperial grade NMR tube (Wilmad). All experiments were recorded at both 298 K and 303 K on a Bruker DRX500 spectrometer equipped with a z-shielded gradient triple resonance probe using standard procedures. Two dimensional nuclear Overhauser and exchange spectroscopy (NOESY) and TOCSY experiments, with mixing times of 200 and 72.5 ms, respectively, were collected with 256 and 1,024 pairs of complex points and acquisition times of 26 and 102 ms in the indirectly and directly acquired dimensions, respectively. Data processing and analysis were carried out on a Silicon Graphics O2 workstation using the programs AZARA and CCPNmr Analysis.

Received 24 March; accepted 8 August 2005.

- Lester, H. A., Dibas, M. I., Dahan, D. S., Leite, J. F. & Dougherty, D. A. Cys-loop receptors: new twists and turns. *Trends Neurosci.* **27**, 329–336 (2004).
- Meeus, N. & Lummis, S. C. R. Proline 307 in the mouse 5-HT_{3A} receptor links binding and function. *pa2online* **1**, 015P (2003).
- Lummis, S. C. R. The transmembrane domain of the 5-HT₃ receptor: its role in selectivity and gating. *Biochem. Soc. Trans.* **32**, 535–539 (2004).
- Bera, A. K., Chatav, M. & Akabas, M. H. GABA_A receptor M2–M3 loop secondary structure and changes in accessibility during channel gating. *J. Biol. Chem.* **277**, 43002–43010 (2002).
- Grosman, C., Salamone, F. N., Sine, S. M. & Auerbach, A. The extracellular linker of muscle acetylcholine receptor channels is a gating control element. *J. Gen. Phys.* **116**, 327–339 (2000).
- Lynch, J. W., Han, N. L. R., Haddrell, J., Pierce, K. D. & Schofield, P. R. The surface accessibility of the glycine receptor M2–M3 loop is increased in the channel open state. *J. Neurosci.* **21**, 2589–2599 (2001).
- Kash, T. L., Jenkins, A., Kelley, J. C., Trudell, J. R. & Harrison, N. L. Coupling of agonist binding to channel gating in the GABA_A receptor. *Nature* **421**, 272–275 (2003).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Nowak, M. W. *et al.* In vivo incorporation of unnatural amino acids into ion channels in *Xenopus* oocyte expression system. *Methods Enzymol.* **293**, 504–529 (1998).
- MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397–412 (1991).
- Dang, H., England, P. M., Farivar, S. S., Dougherty, D. A. & Lester, H. A. Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine(3) receptor gating. *Mol. Pharm.* **57**, 1114–1122 (2000).
- Jabs, A., Weiss, M. S. & Hilgenfeld, R. Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**, 291–304 (1999).
- Dugave, C. & Demange, L. *Cis-trans* isomerization of organic molecules and biomolecules: Implications and applications. *Chem. Rev.* **103**, 2475–2532 (2003).
- Colquhoun, D. Binding, gating, affinity and efficacy: the interpretation of structure-activity relationships for agonists and of the effects of mutating receptors. *Br. J. Pharmacol.* **125**, 924–947 (1998).
- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
- Taly, A. *et al.* Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J.* **88**, 3954–3965 (2005).

17. Zhong, W. *et al.* From *ab initio* quantum mechanics to molecular neurobiology: A cation- π binding site in the nicotinic receptor. *Proc. Natl Acad. Sci. USA* **95**, 12088–12093 (1998).
18. Beene, D. L. *et al.* Cation- π interactions in ligand recognition by serotonergic (5-HT_{3A}) and nicotinic acetylcholine receptors: The anomalous binding properties of nicotine. *Biochemistry* **41**, 10262–10269 (2002).
19. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
20. Gao, F. *et al.* Agonist-mediated conformational changes in acetylcholine-binding protein revealed by simulation and intrinsic tryptophan fluorescence. *J. Biol. Chem.* **280**, 8443–8451 (2005).
21. An, S. S. A. *et al.* Retention of the cis proline conformation in tripeptide fragments of bovine pancreatic ribonuclease A containing a non-natural proline analogue, 5,5-dimethylproline. *J. Am. Chem. Soc.* **121**, 11558–11566 (1999).
22. Kern, D., Schutkowski, M. & Drakenberg, T. Rotational barriers of cis/trans isomerization of proline analogues and their catalysis by cyclophilin. *J. Am. Chem. Soc.* **119**, 8403–8408 (1997).
23. Halab, L. & Lubell, W. D. Use of steric interactions to control peptide turn geometry. Synthesis of type VI β -turn mimics with 5-tert-Butylproline. *J. Org. Chem.* **64**, 3312–3321 (1999).
24. Breiting, H. G., Geetha, N. & Hess, G. P. Inhibition of the serotonin 5-HT₃ receptor by nicotine, cocaine, and fluoxetine investigated by rapid chemical kinetic techniques. *Biochemistry* **40**, 8419–8429 (2001).
25. Kelley, S. P., Dunlop, J. I., Kirkness, E. F., Lambert, J. J. & Peters, J. A. A cytoplasmic region determines single-channel conductance in 5-HT₃ receptors. *Nature* **424**, 321–324 (2003).
26. Cox, C. & Lectka, T. Intramolecular catalysis of amide isomerization: Kinetic consequences of the 5-NH- $\cdots$ N₃ hydrogen bond in prolyl peptides. *J. Am. Chem. Soc.* **120**, 10660–10668 (1998).
27. Brandl, C. J. & Deber, C. M. Hypothesis about the function of membrane-buried proline residues in transport proteins. *Proc. Natl Acad. Sci. USA* **83**, 917–921 (1986).
28. Sansom, M. S. P. & Weinstein, H. Hinges, swivels and switches: the role of prolines in signalling via transmembrane α -helices. *Trends Pharm. Sci.* **21**, 445–451 (2000).
29. Beene, D. L., Price, K. L., Lester, H. A., Dougherty, D. A. & Lummis, S. C. R. L. Tyrosine residues that control binding and gating in the 5-hydroxytryptamine₃ receptor revealed by unnatural amino acid mutagenesis. *J. Neurosci.* **24**, 9097–9104 (2004).
30. Reeves, D. C., Sayed, M. R. F., Chau, P. L., Price, K. L. & Lummis, S. C. R. Prediction of 5-HT₃ receptor agonist-binding residues using homology modeling. *Biophys. J.* **84**, 2338–2344 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.C.R.L. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science. The work at Caltech was supported by the National Institutes of Health. We thank K. L. Price and A. J. Thompson for assistance with modelling.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.D. (dadougherty@caltech.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

The hands that guide

When journalists want to be left alone, they say: "I'm on a deadline." When scientists prefer not to be disturbed, they can use the handy excuse: "I'm writing a grant." According to members of the Graduate Students' Association at the University of California, San Francisco, such excuses aren't in the vocabulary of Renee Reijo Pera. Co-director of the university's programme in human embryonic stem-cell biology, Reijo Pera this year was given the student association's Outstanding Faculty Mentorship Award.

Reijo Pera provides a great start for the new Mentors & Protégés item in *Naturejobs* (see page 254) because she meets my three criteria for excellence in mentoring. She takes time to help protégés with scientific skills — even when she's busy (which is always). She nurtures her students' off-the-bench science skills, such as how to give presentations. And she gives the students long-term career advice.

We suspect that there are many more scientists like Reijo Pera out there who have been recognized for their mentoring contributions, and even more who haven't. But we'd like to begin with the known quantities, and provide them with wider recognition, for three main reasons. First,

because we believe that great mentors should receive credit beyond their home institution or professional organization. Second, recognizing the best mentoring practices sets standards for the field and gives established scientists something to shoot for beyond a publications list. And third, highlighting outstanding supervisors gives prospective postdocs and graduate students something to look for when choosing their own principal investigators.

You can help us in our mission to highlight outstanding mentors by sending us candidates. All nominations should have already received recognition from an organization — be it student, postdoc, institution or professional. Just tell us why the mentor won an award, and why they deserve wider recognition. Please send your nominations to naturejobseditor@naturedc.com.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 345 Park Avenue South,
 10th Floor, New York,

NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Steven Williams, president, Wildlife Management Institute, Washington DC



2002-05: Director, US Fish and Wildlife Service
1995-2001: Secretary, Kansas Department of Wildlife and Parks, Topeka
1992-95: Deputy executive director, Pennsylvania Game Commission, Harrisburg
1989-92: Assistant director for wildlife, Massachusetts Division of Fisheries and Wildlife, Westboro

Steven Williams has been a rock-skippling, wilderness-rambling, wildlife-ogling guy for as long as he can remember. Not surprisingly, he turned those interests into degrees in biology and resource management. His path, however, has gone from the wilderness to the halls of US national government and back again.

Wanting to focus on applied management following his PhD in Forest Resources at Pennsylvania State University, Williams began working as a wildlife biologist at the Massachusetts Division of Fisheries and Wildlife. He continued with three different state agencies for the next 17 years, moving his way up management. He got his first taste of the politics of civil service in 1995, when he lost his job as deputy executive director of the Pennsylvania Game Commission for what he describes as political reasons. It was his greatest challenge. "At that time, I was really questioning whether to continue in this profession," he says. Following advice from mentors, he decided to keep going. And he's glad he did. "The experience opened the door to the two best career experiences I've had in my life," he says.

Williams became the secretary of the Kansas Department of Wildlife and Parks. Noticed for his ability to bring different stakeholders together to resolve natural-resource conflicts, he was asked to become director of the US Fish and Wildlife Service. Although he loved the position, in which he spent three years, the red tape in Washington DC often proved a hurdle.

He left the federal post earlier this year to take up the position of president of the Wildlife Management Institute (WMI) based in Washington DC. Being at a non-profit organization dedicated to science-based management allows him to focus more on conservation, he says. At the Fish and Wildlife Service, he was pulled in many directions. But the WMI's mission is more centred on traditional conservation issues such as habitat and population management.

Williams says that the best advice he can give young scientists, other than finding a good spouse, is to "set yourself apart" by combining skill sets in a unique way. In gaining a background in statistics and the then nascent field of geographic information systems, he says, he offered more value to his future employers. In fact, his first employer has since told him that they hired him to see how those skills could help them. He has proved that the chance was worth taking. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Approachable award-winner

This year, the Graduate Students' Association at the University of California, San Francisco, honoured my adviser, stem-cell biologist Renee Reijo Pera, with its annual Outstanding Faculty Mentorship Award. I was one of five current or former students who wrote nomination letters filled with examples of how Renee consistently praises our strengths, shores up our weaknesses, and shows real concern for our career development and personal well-being.

One of Renee's strengths is her approachability. Every student in our lab has a regular weekly meeting with Renee, no matter how full her schedule is. Even when she's out of town, she stays in close e-mail contact. I remember once when Renee was rushing to finish writing a huge grant proposal, she still made time at the last minute to go over slides for a presentation I was giving the next day. Renee's availability and approachability make us feel comfortable in the lab and keep our projects focused and on track.

Renee also prepares us for a career in science by repeatedly encouraging us to present posters, give talks, take relevant classes and write papers. She lets us work independently, but provides help when needed. She also helps us set up collaborations with faculty members from other departments and

institutions, so that we can develop our own network of contacts. As we approach graduation, Renee has spent hours talking with us about different career options, what to look for in an employer, and how best to present ourselves.

Renee encourages us to make time for personal or family issues and extracurricular activities, and doesn't concern herself with our 'face time' in the lab. Many of us have been active in student organizations on campus, and she has often volunteered for activities when we needed involvement from faculty members. She understands that students who lead balanced lives come to work ready to tackle experiments with energy and a clear mind.

Simply put, Renee is a good mentor because she cares enough to put the time and effort in to helping us succeed no matter what we endeavour — a trait that is, unfortunately, not common enough.

At the University of California, San Francisco, we are organizing a mentorship course for faculty members and students. We hope that these courses and role models such as Renee will help to improve the training of the next generation of scientists. ■

Joyce Tung is a graduate student at the University of California, San Francisco.

GRADUATE JOURNAL

Post-holiday revelations

I have been back in the laboratory for two weeks after a break between finishing my master's and starting my doctorate. Coming back, I realized how much I had missed thinking about scientific problems during my holiday, as well as planning experiments and reading papers. There is no doubt: research now belongs in my life.

The other thing I immediately sensed is how tedious everyday lab life can be. I have to prepare all the materials for my experiments. If a reagent runs out, I have to make a note, order and restock. If I don't, the experiment will not happen. Organizational skills are very important. This part of the job may not be the most scientifically interesting, but I see now that there is a benefit.

I always wondered why scientists were primary targets for headhunters scouting for major consulting companies. I would never drop out of science to wear a suit and advise business people every day. But I realize how problem analysis, time and resource management skills, and the ability to work independently — so crucial in the lab — are also a fantastic set of qualities for any job type.

But I'm still not tempted to sell my skills to private industry. I have three more years to go before I finish my doctorate. Plenty to learn, plenty to organize. ■

Tobias Langenhan is a graduate student in neuroscience at the University of Oxford, UK.

The crime of the century

A little family planning.

Geoff Brumfiel

A cold spring rain pounded the windshield of her cruiser. She sighed, turned up the collar on her overcoat and sprinted from the car to the porch. It was too early to play detective, but here she was.

"Hey Joe," she mumbled, flashing her shield to the cop outside the front door. She stepped inside and lit up a cigarette, pausing to let her eyes adjust to the gloom.

"Come on in, lieutenant," shouted a voice from the next room.

"Christ," she thought to herself. "How does he always beat me to the scene? And why is he sounding so goddamn perky?" She checked her watch, it wasn't even seven.

The body was curled up on the kitchen floor like a baby; the rookie stooped over it, peering at it like he'd never seen a corpse before. Who knows, maybe he hadn't; he'd only been on the force for 25 years.

He glanced up and flashed her a grin. "Morning lieutenant."

She glared at him. "What can you tell me?"

"Well," he said, unplugging his tablet from the data port on the dead man's neck. "His name is Henry Watson, age 275, Washington DC resident. The guy is — sorry, was — a researcher at the National Institute of Radical Life Extension up in Bethesda. Looks like he worked on organ regeneration, mainly lungs." He paused. "I guess he's the reason you can puff on those cancer sticks without getting cancer."

"Spare me the poor schmuck's life story," she snapped. "What'd he die of?"

"Looks like a myocardial infarction," he said, glancing down at his tablet.

"Heart attack, huh?" she smirked. "Haven't heard that one for a while."

She flicked her cigarette into the sink and walked over to the dead man's refrigerator. Pizza, soda, cake. "This guy ate like a pig," she thought to herself. Not that it mattered what you ate these days.

She took a soda, and walked back to the body to have a look. The rookie interrupted her thoughts: "Our boys think he's been dead for at least a week, but it's hard to tell. His immunobots are still fighting off infections, so he hasn't started stinking the place up," he said.

"I see," she said, and pulled from her pocket a thin, silver tablet identical to her partner's. She pushed a button on



JACEY

the side, and it whirled into life.

"Good morning lieutenant," the tablet said smoothly.

"Morning Gracie," she replied. "I need you to run a diagnosis of our stiff's nanosystems."

"One moment, please," the tablet hummed softly in her hand. "I'm showing an abnormality. Mr Watson's cardiobots have been inactive since 2246."

"That's almost 15 years!" the rookie said.

"Yeah," the lieutenant growled. "And with his diet, it's no wonder he dropped dead. Without his little nanofriends to scrub his arteries clean, the guy's cholesterol must have gone through the roof. Gracie, do a round-up of the usual suspects."

"Hold please, processing," the tablet droned. "Wait a moment, lieutenant, I have a suspect. Mrs Erin Marie Howard, age 72, of 414 Darrow Street, Ann Arbor, Michigan. Mrs Howard is Mr Watson's great-great-great-granddaughter. Married to John Howard, age 94. She's currently employed as a biomedical nanotechnician at GSK-Syngenta-PhilipMorris. I'm showing she illegally deactivated Mr Watson's cardiobots from a remote terminal on March 10, 2246 at 14:24:37."

"But why would she kill her great-great-grandfather?" the rookie asked.

"That's three greats," the lieutenant interrupted.

"I'm showing Mrs Howard has a conception licence application pending with the Bureau of Births and Population Management," the tablet crooned.

"Well there's your answer," the lieutenant said. "The girl wanted to have herself a baby, and she knew the feds wouldn't let her unless someone in the family happened to keel over, which doesn't happen too often these days. Gracie, ask Judge Hastings to issue us an arrest warrant."

"Wow," said the rookie. "I've never heard of a case like that. Talk about cold-blooded."

"I'd say something like this comes along every hundred years or so," the lieutenant sighed and leaned against the kitchen counter, taking a swig of the deceased's soda. "The girl probably barely knew Watson. She was so much younger than he was; they're hardly even related. Why not bump the old guy off? Clever too, giving him 15 years to do it to himself — too bad for her our boys used the time to develop AI crimepads like Gracie here."

"Judge Hastings would like to know what charges you're filing against Mrs Howard," the tablet asked.

"Murder, Gracie," the lieutenant replied. "Murder by natural causes."

Geoff Brumfiel lives and works in Washington as *Nature's* physical sciences correspondent.